January 2015

Calcium Requirement Distribution Via Bone Growth Modeling

Michael Ryan Lawlor
Purdue University

Graduate School Form 30
Updated 1/15/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Michael Ryan Lawlor

Entitled

Calcium Requirement Distribution Via Bone Growth Modeling

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

George McCabe
Chair

Bruce Craig

Connie Weaver

Shengchun Kong

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): George McCabe

Approved by: Hao Zhang 9/24/2015

Head of the Departmental Graduate Program Date

CALCIUM REQUIREMENT DISTRIBUTION VIA BONE GROWTH MODELING

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Michael R. Lawlor

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2015

Purdue University

West Lafayette, Indiana

To Brian Lawlor, Joan Zupan, Marty Lawlor, and Gayle Lawlor.

ACKNOWLEDGMENTS

I need to convey my most heartfelt appreciation for my advisor, Professor George P. McCabe. His patience and encouragement helped keep me focused and motivated during the research process. He kept me in the program and made it possible to write this thesis. His suggestions improved the quality of the work.

A special thanks to Professor Donald Bailey et al. for providing the Saskatchewan data set discussed within.

I would also like to thank Professor Bruce Craig, Professor Shengchun Kong, and Professor Connie Weaver for serving on my committee. Their suggestions and critiques were important. I would like to thank Ellen Gundlach for her guidance during my years as a teaching assistant. The secretarial and computer system staffs deserve credit for their help through the years.

I would like to thank Professor Rebecca Doerge for her guidance, understanding, and direction during my time as a Ph.D. candidate.

A special thanks to Professor Peter Thompson for his guidance while at Wabash. He helped me choose a life of academia, particularly at Purdue University.

Lastly, I would like to thank my family, especially my brother, for always being there to support me. Together with our parents, their love and belief in me, has helped me maintain a balance between life and school.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Figure Page

ABBREVIATIONS

TBBMC	total body bone mineral content
TBBMCV	total body bone mineral content velocity
PBMCV	peak bone mineral content velocity
RDA	recommended dietary allowance
EAR	estimated average requirement
DRI	dietery reference intakes
AI	adequate intake
MVN	multivariate normality

GLOSSARY

TBBMC	total amount of bone mineral in the body, usually in kg
TBBMCV	change in TBBMC over time, usually in kg/year
PBMCV	the largest value of TBBMCV for an individual

ABSTRACT

Lawlor, Michael R. PhD, Purdue University, December 2015. Calcium Requirement Distribution via Bone Growth Modeling. Major Professor: George P. McCabe.

This dissertation focuses on estimating calcium requirements using bone growth data in adolescents. We estimate the peak rate of growth of bone and the variability of this peak growth rate to provide Dietary Reference Intakes (DRIs) for calcium. The specific DRIs of interest are Estimated Average Requirement (EAR) and Recommended Dietary Allowance (RDA). Included within the analysis is the age at which this peak occurs. A generalized logistic curve uses age as the explanatory variable to predict total body bone mineral content (TBBMC), which is the mass of the mineral content within the bones in the entire body. The population of subjects of interest is males and females aged 9 through 18. We estimate the distribution of these growth curves across individuals. We assume that this distribution is specified by parameters that are jointly Normal. Parameters of these models are estimated using the Saskatchewan data set from Bailey et al. The subsequent DRI analysis uses values at the peak rate of growth as well as an average rate of growth over the current age guidelines within the factorial method for both males and females. Additionally, the DRI analysis compares the use of a factorial method versus different models to relate the amount of calcium retained to the amount of calcium intake required. Future work will investigate the incorporation of race, body mass index, height, weight, and other anthropomorphic variables. For this portion, material will come from the Saskatchewan, Iowa, and Camp Calcium data sets.

1. INTRODUCTION

1.1 Calcium Requirements and Bailey's Papers

The main goal of this research is to estimate calcium requirements for children aged 9 to 18. This chapter discusses the current requirements, their formulation, and an idea proposed by Bailey et al. that provided a different manner of thinking with respect to their calculation.

1.1.1 Dietary Reference Intakes

To begin, it is necessary to define terms associated with dietary reference intakes (DRIs). Values of DRIs are based on the concept that requirements have a distribution for individuals within a population. The two main DRI values are estimated average requirement (EAR) and recommended daily allowance (RDA). The EAR is the daily intake required to meet the needs of 50 percent of the individuals in a life stage or gender group. [1, p.23] Thus it is the median of the requirement distribution and it would also be the mean if the requirement distribution is symmetric. The RDA is the daily intake value associated with the 97.5^{th} percentile of the requirement distribution. The following is an excerpt from the Institute of Medicine (1997): [1, p.23-4]

> *The RDA is intended primarily for use as a goal for daily intake by individuals. The EAR forms the basis for setting the RDA. If the variation in requirements is well defined and the requirement is normally distributed, the RDA is set at 2 standard deviations (SD) above the EAR:*
>
> $$RDA = EAR + 2SD_{EAR}.$$

*If the SDs reported in studies are inconsistent, or if sufficient data on variation in requirements are not available for other reasons, a standard estimate of variance will be applied. This estimate assumes a coefficient of variation (CV; SD divided by the mean * 100) of 10 percent, which is equal to 1 SD, such that*

$$RDA = 1.2 * EAR.$$

Using a Normal Distribution, the 2 in the first formula is because the 97.5^{th} percentile is about 2 SDs above the mean, which in this case is the EAR since the median is also the mean. In the second formula, SD = .1 EAR, so

$$EAR + 2SD = EAR + 2 * .1EAR = 1.2EAR.$$

1.1.2 Factorial Method

This information has been used in conjunction with analysis done by Vatanparast et al. [2] in the 2011 "Dietary Reference Intakes for Calcium and Vitamin D". [3] Vatanparast et al. calculated average calcium accretion in milligrams per day (mg/d) and applied this using the factorial method to find the associated calcium intake. The results are displayed within Figure 1.1.

The estimated total intake (adjusted for absorption) column was used to calculate the EARs for each age group. The two groups that are for children ages 9-18 are formed by averaging the 9-13 and 14-18 year groups' average calcium accretion in mg/d for that gender. For example, the average calcium accretion in mg/d for 9-18 Female = 121 = $\frac{151+92}{2}$, where the 151 is the value for the 9-13 year old females and the 92 is the value for the 14-18 year old females. The total need in mg/d column is the sum of the first four columns. The first four columns represent accretion (growth), urinary losses, fecal losses, and sweat losses. This total is then divided by .38 (an

Figure 1.1.: Factorial Method

TABLE 5-2 Calcium Intake Estimated to Achieve Average Bone Calcium Accretion for Children and Adolescents Using the Factorial Method

Study Author, Year	Age/ Gender	Average Calcium Accretion (mg/day)	Urinary Losses (mg/day)	Endogenous Fecal Calcium Losses (mg/day)	Sweat Losses (mg/day)	Total Needed (mg/day)	Absorption (percent)	Estimated Total Intake (Adjusted for Absorption)
Lynch et al., 2007	1–3 Male/Female	142	34	40	—	216	45.6	474
Abrams et al., 1999; Ames et al., 1999	4–8 Male/Female	140–160	40	50	—	240	30.0	800
Vatanparast et al., 2010	9–13 Female	151	106	112	55	424	38.0	1,116
	9–13 Male	141	127	108	55	465	38.0	1,224
	14–18 Female	92	106	112	55	365	38.0	961
	14–18 Male	210	127	105	55	500	38.0	1,316
	9–18 Female	121	106	112	55	394	38.0	1,037
	9–18 Male	175	127	108	55	465	38.0	1,224

adjustment made for absorption rate) to get the final values of estimated total intake (adjusted for absorption) in the last column.

> *Thus, for reference values for both males and females in the 9- to 13- and 14- to 18-year life stages, the differences in calcium intake to achieve mean bone calcium accretion as elucidated by Vatanparast et al. (2010) have been interpolated between 9- to 18-year old females (1,037) and males (1,224). This interpolation yields an estimated mean need for calcium for males and females of 1,100 mg/d with rounding, a value approximately at the midpoint between the two groups. Again, assuming a normal distribution, this estimate to achieve a mean calcium accretion represents the median, and, thus, an EAR. The EAR is therefore set at 1,100 mg for both males and females for both life stages encompassed by the 9 through 18 year age range. In order to cover 97.5 percent of the population, an estimated RDA value for calcium of 1,300 mg/d is established. [3, p.354]*

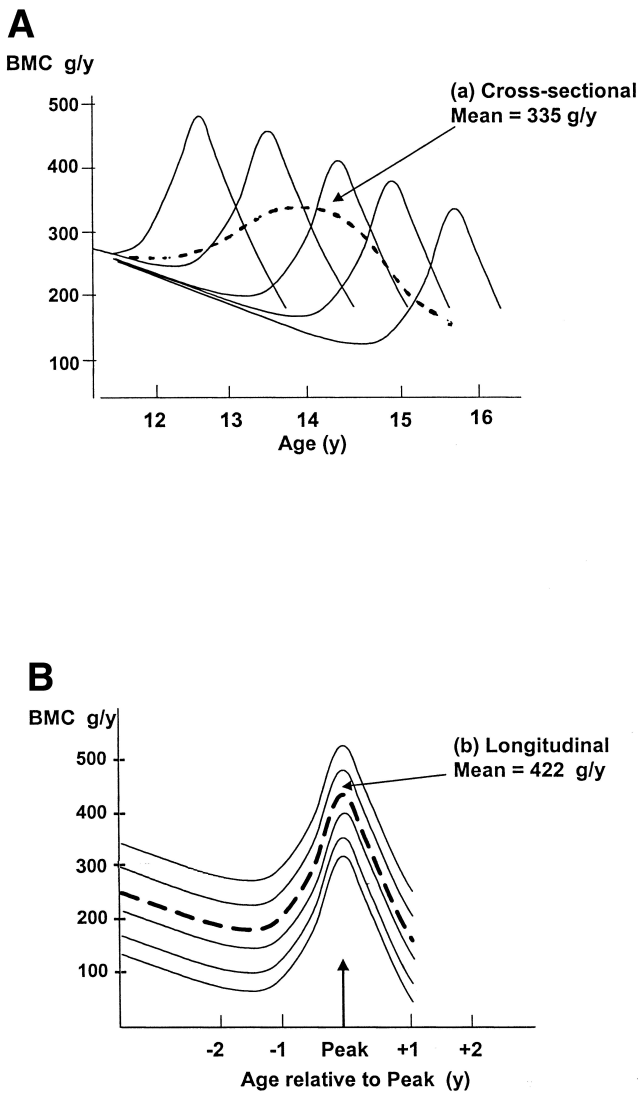
1.1.3 Growth of Bone

Bailey et al.(2000) study Total Body Bone Mineral Content (TBBMC) in grams (g) and TBBMC Velocity (TBBMCV), or change in TBBMC, in grams/year (g/y).

TBBMC is the total mass of the minerals in bones in the body and TBBMCV is the change in this mass over time (usually per year). They focused only on children that exhibited clear peaks in height and TBBMC velocities. Their data consisted of 113 males and 115 females. Their results pertain to a subset of these data consisting of 60 males and 53 females. A statement was made that this mainly eliminated children whose age of peak height velocity (PHV) was before their age at the time of first measurement. For each child, they fit cubic splines to the height and TBBMC velocity curves. From this, they calculated means and SDs of the age and value of PHV and Peak Total Body Bone Mineral Content Velocity (PBMCV), which is the mineral growth of the bone. Since these curves related to individuals, they asked whether they should be pooled chronologically by age (a cross-sectional approach) or aligned by peak and then averaged. They lined up TBBMCV curves by peak and used TBBMCV averaged over age as a basis of comparison. They showed that a cross-sectional view underestimated the calcium retention at peak. [4] Figure 1.3 illustrates the basic idea.

In Figure 1.2, the curves represent PBMCV for individual males. In both graphs, the y-axis represents TBBMCV in g/y, which is the rate of bone accrual in the body. For graph A, the x-axis is age in years. For graph B, the x-axis is age relative to peak in years. For graph B, each individual was aligned according to the age of their peak regardless of whether that occurred at 13 or 16 and this peak was placed in the center of the graph. In both graphs, the dashed line represents the average curve of the 5 individuals. In graph A, the dashed curve has a peak at around 14 years of age and a value of 335 g/y. In graph B, the peak value is about 422 g/y. These values were calculated over the 5 individuals plotted. The females have peaks at roughly 240 and 322 g/y for graphs A and B respectively. The values for all males 320 and 407 g/y for graphs A and B respectively.

Figure 1.2.: Cross-Sectional vs. Lined-up Peaks from Bailey et al. (2000)



In their paper, Bailey et al. focused on the peak aligned values, 407 g/y and 322 g/y for males and females respectively. This approach assumes that when the growth of bone is most, the need for calcium is also the greatest, and calcium requirements should reflect this peak, not the average. These peaks will be the focus for this next part.

These values at the peaks are not the quantity of interest, rather they are a means to calculate a calcium requirement. In order to do so, they must first be translated into the proper units, in this case mg/d of calcium. The conversion for this is $\frac{1000}{365}.322$. The 1,000 converts from g to mg, the 365 converts from years to days, and the .322 is the conversion factor relating bone mass to calcium mass. The peaks for males and females then become 359 and 284 mg/d of calcium. Now that the numbers represent calcium retained, they can be used within the factorial method.

The factorial method is a way of relating intake, retention, and losses for calcium. Losses are split into 3 parts: urinary losses, endogenous fecal losses, and sweat losses. The idea is that retention is a function of intake, losses, and absorption rate. In general form,

$$retention = intake * absorption - losses.$$

In this context, retention is known and predicted intake is desired. Therefore, the general formula of interest is:

$$intake = \frac{retention \quad + \quad losses}{absorption}.$$

The factorial method [1], adjusts peak retention values for calcium losses and absorption efficiency. The end results are peak intake values of 1,708 and 1,466 mg/d of calcium for males and females respectively. A summary of this information is in Table 1.1.

Table 1.1.: Peak Calcium Intake from Bailey et al. (2000)

	PBMCV in g/y	Ca Intake needed in mg/d
Males	407	1708
Females	322	1466

It is important to keep in mind that these are just the average values; there was no addition of a factor pertaining to the standard deviation of the distribution. Therefore, they would correspond to the EAR not the RDA. Both calcium intake values are above the EAR of 1,100 mg/d.

Bailey et al.used cubic splines to model TBBMCV. Within their analyses they found ages and values of PBMCV. For the males, age of PMBCV had a mean and standard deviation of 14.0 y and 1.0 y respectively. For the females, age of PBMCV had a mean and standard deviation of 12.5 y and .9 y. While they calculated means and SDs with their model, there was not an established distribution for PBMCV. We wanted to model TBBMC in a way that would not only estimate the ages and values of PBMCV, but also include a distribution for PBMCV. It is important to be able to establish this distribution to estimate an RDA. (Recall RDA is the 97.5^{th} percentile of the requirement distribution.)

Table 1.2 is an extension of table 3 in Bailey et al.(2010). [5] The extension was to find these values for each year, not just the two age categories, 9-13 and 14-18. The calcium retention values are directly from table 2 in this same paper. They represent the amount of calcium retained in the bone in mg/d. The calcium intake values are the results of the factorial method using the calcium retained values as inputs and represent the calcium needed (in mg/d) to have that amount of calcium retention. As an example, for 9 year old males, $1077 = \frac{119+290}{.38}$. The value of 290 is the total value for losses for males in mg/d (this is 273 for females). The value of .38 represents an absorption rate of 38%. These are standard values from the IOM. [3] Figure 1.4 is a plot of calcium needs (the intakes in Table 1.2) and EAR for calcium by age.

Figure 1.3 is a plot of the calcium intake values form Table 1.2 according to gender. In Figure 1.3, both males and females intake requirements are increasing to an age (13 for females and 14 for males) and then decrease afterwards. Another general

Table 1.2.: Extension of Table 3 in Bailey et al. (2010). All Ca variables are in mg/d.

	Male		Female	
Age	Ca Retained	Ca Intake	Ca Retained	Ca Intake
9	119	1077	88	949
10	101	1028	99	980
11	128	1099	145	1099
12	154	1169	190	1218
13	204	1301	235	1336
14	296	1543	164	1150
15	262	1452	107	1001
16	236	1384	67	895
17	143	1140	50	849
18	111	1056	74	914

pattern is that the female requirements are lower at the beginning, pass the males during the female's peak time (12-13), and then go back below the male level after their peak. The calcium intake values do not include an added standard deviation, so they are comparable to EAR.

Looking at this plot, EAR is underestimating the calcium need around the time of peak growth and need. For females it is underestimating from about 11 to 14; whereas this time of underestimation is about 11 to 17 in males.

Figures 1.5 - 1.8 focus on the density curves for TBBMCV, which is the difference in consecutive TBBMC values. The age groups for requirements are 9 to 13 and 14 to 18. Since the difference is needed, the first group includes 8 year olds too. The x-axis is TBBMC in g. The y-axis is the probability for the density curves. Figure 1.5 is a graphical representation of Table 1 in Bailey et al.(2010) for the males based on ages. The standard deviations for the 8-13 year olds increase with age. However, the

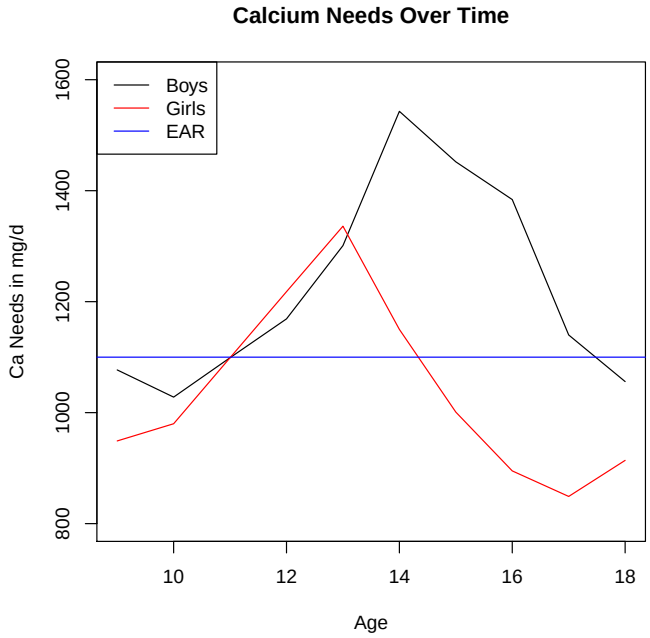


Figure 1.3.: This is the calcium needs adjusted for absorption (using the factorial method) for males and females respectively with Estimated Average Requirement (EAR).

standard deviations for the 14-18 year olds are very similar, just the means increase with age. Figure 1.6 shows the density curves for the females based on ages. They have the same color scheme as the male densities in Figure 1.5.

Similar to the males, the female densities exhibit increasing means with age. They also show increasing standard deviations with age, but only for the 8-13 year old group, not the 14-18 year old group.

The values for the means and SDs for Figures 1.5 and 1.6 came from Table 1.1 in Bailey et al. (2010).

Table 1.3.: Table 1 from Bailey et al. (2010)

			TBBMC (g)			
	n		Males		Females	
Age (years)	Males	Females	Mean	SD	Mean	SD
8	6	18	797.3	72.2	790.5	178.7
9	19	34	932.5	102.6	889.8	225.8
10	32	53	1046.5	148	1002.3	234.5
11	53	65	1191.1	169.6	1166.2	297.6
12	75	78	1365.9	224.3	1381.2	322.7
13	88	92	1597.6	307.4	1647.2	342.6
14	89	95	1933.5	399.2	1833.2	324.9
15	79	87	2230.1	409.5	1954.9	267.1
16	66	61	2497.4	373.2	2030.8	272.3
17	51	45	2659.6	391.7	2086.8	293.2
18	36	34	2785.5	424.2	2171.2	339.0

Figure 1.4.: Normal Distributions for males based on individual ages

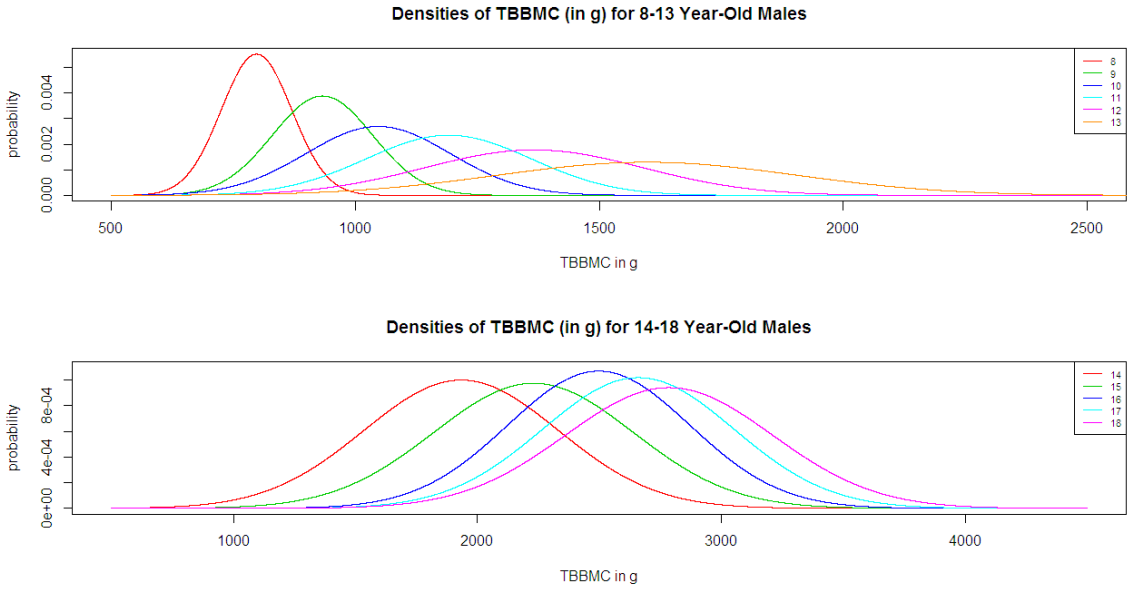
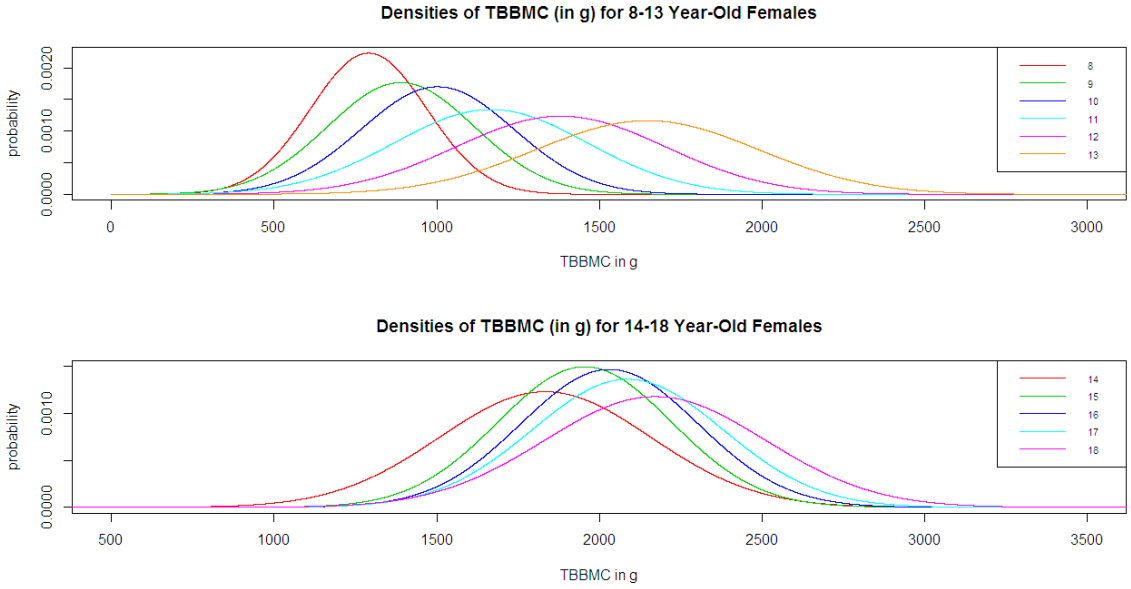


Figure 1.5.: Normal Distributions for females based on individual ages



1.2 Mixture Distributions

The next part focuses on mixture distributions and densities. A mixture represents a weighted sum of a group of underlying random variables. The weights

Figure 1.6.: Table 1 Relating to TBBMC from Bailey et al. (2010)

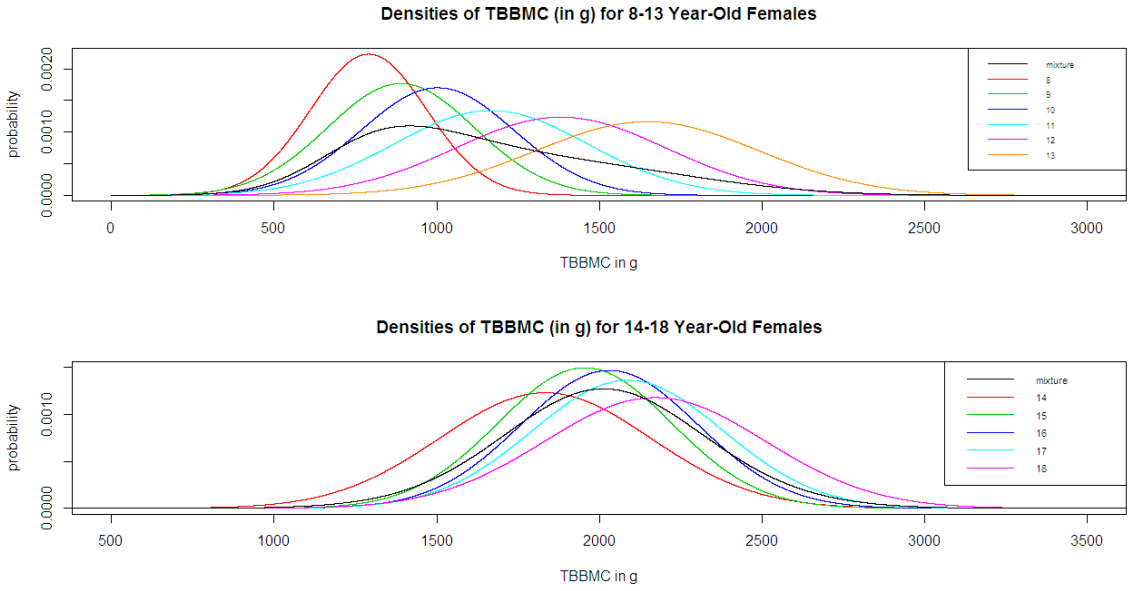
			TBBMC (g)			
	n		B		G	
Age (years)	B	G	Mean	SD	Mean	SD
8	6	18	797·3	72·2	790·5	178·7
9	19	34	932·5	102·6	889·8	225·8
10	32	53	1046·5	148·0	1002·3	234·5
11	53	65	1191·1	169·6	1166·2	297·6
12	75	78	1365·9	224·3	1381·2	322·7
13	88	92	1597·6	307·4	1647·2	342·6
14	89	95	1933·5	399·2	1833·2	324·9
15	79	87	2230·1	409·5	1954·9	267·1
16	66	61	2497·4	373·2	2030·8	272·3
17	51	45	2659·6	391·7	2086·8	293·2
18	36	34	2785·5	424·2	2171·2	339·0

are probabilities of the mixture coming from that particular random variable. The weights need to follow the rules of a probability distribution. Namely, they all need to be between 0 and 1 and they need to sum to 1. A mixture density is a weighted sum of the densities of the underlying random variables. The purpose of this section is to see if applying mixture densities to the ages in each requirement group would yield a useful distribution. One assumption made is that the individual distributions are Normal and the goal is to see if the mixtures, which would represent the overall requirement groups, would also be Normal.

Figures 1.7 and 1.8 have the same color schemes as Figure 1.5 but with an additional color of black, that is the mixture distribution.

The mixture distributions are weighted sums of the individual distributions. In this case, the weights are equal for each age. In the first group (ages 8-13), there are 6 different ages, so the weights are $\frac{1}{6}$. In the second group (ages 14-18), there are 5 different ages, so the weights are $\frac{1}{5}$. To show how these black lines were calculated in Figures 1.7 and 1.8, let us examine the mixture for males ages 8-13.

Figure 1.7.: Normal Distributions for males based on individual ages with the overall mixture



Let i represent the age, so in this case it will be 8, 9, ... , 13. Let X_i denote the males' TBBMC in g for age i. Let f_i denote the density for X_i. Let M be the weighted sum of the X_i. Then, the mixture density, f(m), is

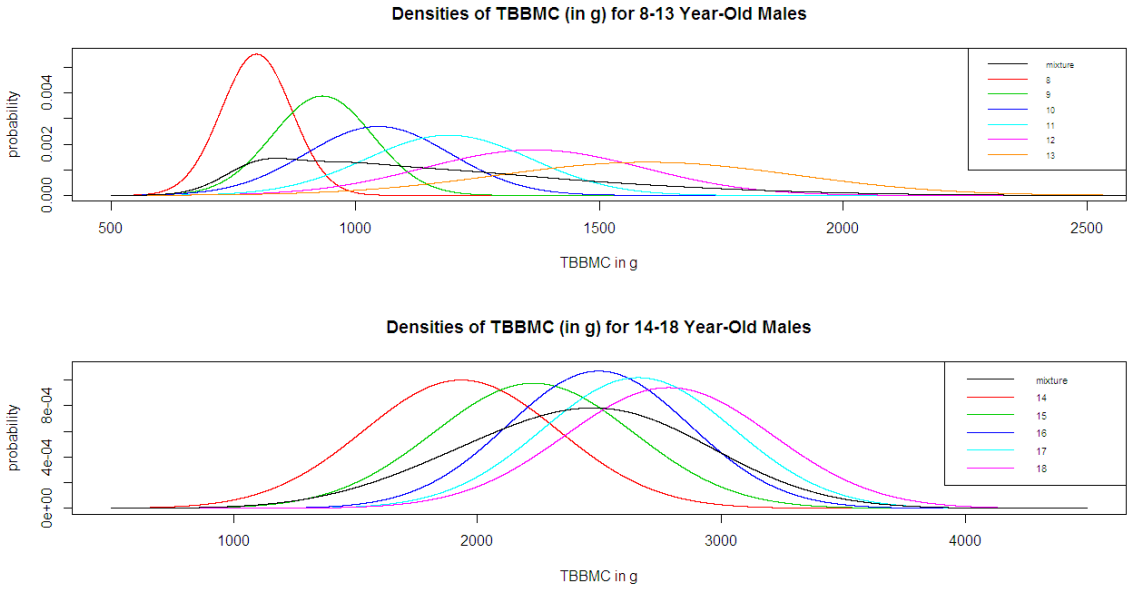
$$f(m) = \sum_{i=8}^{13} p_i f_i(x) = \sum_{i=8}^{13} \frac{1}{6} f_i(x).$$

It is assumed that the $X_i \sim N(\mu_i, \sigma_i^2)$. If one wants the moments of the mixture distribution, they are the weighted sums of the moments of the individual distributions. For example,

$$EM = \sum_{i=8}^{13} \frac{1}{6} \mu_i,$$

$$E[M^2] = \sum_{i=8}^{13} \frac{1}{6} (\mu_i^2 + \sigma_i^2),$$

Figure 1.8.: Normal Distributions for females based on individual ages with the overall mixture



and

$$Var(M) = \sum_{i=8}^{13} \frac{1}{6}(\mu_i^2 + \sigma_i^2) - (\sum_{i=8}^{13} p_i \mu_i)^2.$$

Figures 1.7 and 1.8 were made based on these calculations. Their means and standard deviations are given in Table 1.3. Looking at Table 1.3, the values for the standard deviations are quite large.

Table 1.4.: TBBMC Mixture Distributions

Mixture Distribution	TBBMC Mean (in g)	TBBMC SD (in g)
Males ages 8-13	1155	327
Males ages 14-18	2421	504
Females ages 8-13	1146	402
Females ages 14-18	2015	322

It is interesting to note that the mixture distributions for the 8-13 year old children do not appear to be Normal for either gender. However, the 14 - 18 year old children's mixtures appear more Normal, especially for the females.

1.3 Conclusions

The amount of calcium needed to meet the calcium accretion need is assumed to be Normal, with a mean of 1,100 mg/d. However, there is not a current estimate of the standard deviation of calcium needed, but a coefficient of variation is used in its place in the calculation of the RDA. According to Bailey et al.(2000), the cross-sectional viewpoint underestimates the calcium needed to meet the demands at the time of peak bone growth. Even if the distributions of TBBMCV by age were Normal, the corresponding mixtures would not be.

2. DATA SET FROM BAILEY ET AL.

2.1 Descriptive Statistics

The data set used was from Bailey et al., and is referred to as the Bailey data set. The subjects are adolescents from Saskatchewan. The variables are: Subject ID, Sequence Number, Age at Test in y, Age of Peak Height Velocity (PHV) in y, Height in cm, Weight in kg, Total Body Bone Mineral Content (TBBMC) in g, Age of Peak BMC Velocity (PBMCV) in y, Date of Visit, and Age of Menarche. Descriptive statistics and other characteristics of the variables are the focus of this chapter.

Subject ID is in the 1000s if the subject is a male and in the 2000s if the subject is a female. From this, a dummy variable called SexMale was created. It has a value of 1 if the individual is a male and 0 if the subject is female. Sequence number represents the time of the visit for each individual. The first visit gets a sequence number of 1. Since data was collected (about) every 6 months, the sequence number reflects this time frame. As an example, the subject with the ID number of 1001 had their first 10 visits (about) every 6 months. Additionally, he had a follow-up after his first 10 visits that was 7 years later. This visit has a sequence number of 24 (10 + 2*7). TBBMC is the mass of bone in the entire body (here measured in g). Peak height velocity (PHV) and Peak Bone Mineral Content Velocity (PBMCV) are the maximum values of height and BMC growth. The variables in this data set reflect the age (in y) at which these maximum values take place. Menarche is the first time a female has her menstrual cycle.

The following is a table describing the number of subjects and observations by gender.

Table 2.1.: Number of Subjects and Observations by Gender

	Number of Subjects	Number of Observations
Total	251	3,433
Females	134	1,836
Males	117	1,597

Here is a 5 number summary of the number of visits per subject:

Table 2.2.: Summary for the Number of Visits per Subject

Minimum	First Quartile	Median	Third Quartile	Maximum
1	10.5	15	18	22

In the analysis by Bailey et al., they looked at both TBBMC and Height as response variables. There are some visits that do not have measurements for all the variables. TBBMC was measured on 2,106 of the visits and height was measured on 3,419 of the visits. Next are some characteristics of the patient.

Table 2.3.: Height in cm

Data Set	Min	Q1	Median	Mean	Q3	Max	number NAs
Total	117.0	154.0	163.4	162.5	171.6	193.5	14
Female	117.0	153.6	161.6	159.0	166.8	189.9	10
Male	120.0	154.4	169.6	166.5	179.1	193.5	4

Not surprisingly, the heights of males are higher across the board than those for females. In general, the weight is higher for males than for females. However, the maximum weight of the data set is female.

Table 2.4.: Weight in kg

Data Set	Min	Q1	Median	Mean	Q3	Max	number NAs
Total	19.40	44.10	56.06	57.97	69.60	142.00	15
Female	19.40	43.70	54.00	55.68	65.00	142.00	10
Male	20.90	44.70	59.75	60.60	74.25	116.00	5

Table 2.5.: TBBMC in g

Data Set	Min	Q1	Median	Mean	Q3	Max	Number of NAs
Total	487.1	1532.0	2015.0	2038.0	2506.0	3843.0	1327
Female	487.1	1521.0	1909.0	1846.0	2203.0	3632.0	691
Male	642.7	1551.0	2357.0	2267.0	2911.0	3843.0	636

The males have higher TBBMC values than the females with the differences for the medians, means, and third quartiles being quite substantial.

Table 2.6.: Age of Menarche

Data Set	Min	Q1	Median	Mean	Q3	Max	Number of NAs
Female	9.624	12.050	12.810	12.710	13.310	15.000	36

The values for age of menarche are consistently about a half to a full year behind the ages of Peak Height Velocity (PHV) in females.

Females reached age of PHV younger than males (all across the 5 number summary and mean). Interestingly though, both groups seem to exhibit a range of about 4.5 years for age of PHV.

Again, the females have a lower age to a certain benchmark, in this case PBMCV. These values of age of PBMCV were calculated by Bailey et al. There are 111 total NAs when there are only 251 people in the study. That is a high percentage, about

Table 2.7.: Age of Peak Height Velocity (PHV) summary

Data Set	Min	Q1	Median	Mean	Q3	Max	number NAs
Total	9.17	11.80	12.59	12.60	13.44	15.88	12
Female	9.17	11.31	11.90	11.86	12.45	13.63	6
Male	11.14	12.79	13.45	13.46	14.04	15.88	6

Table 2.8.: Age of Peak BMC Velocity or PBMCV

Data Set	Min	Q1	Median	Mean	Q3	Max	Number of NAs
Total	10.46	12.52	13.41	13.39	14.14	15.93	111
Female	10.46	12.05	12.61	12.67	13.27	14.78	64
Male	12.03	13.57	14.02	14.11	14.87	15.93	47

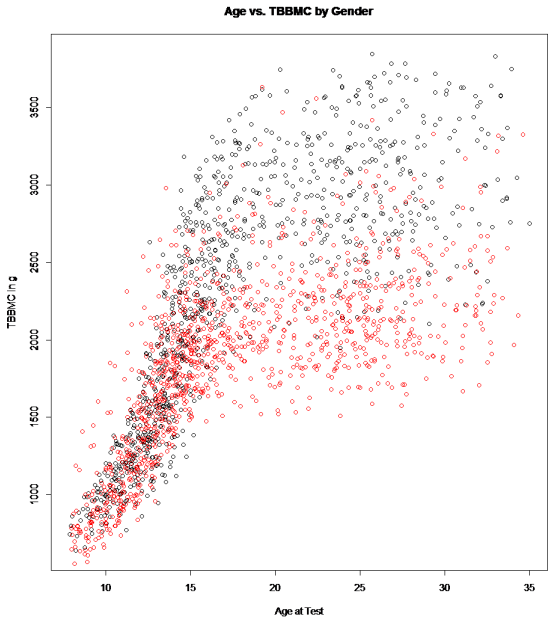
44% (and about 48% for females). They then omitted people without ages of PBMCV from their analysis when computing their estimates. This is one of the advantages of our method. While it may be difficult to fit an individual curve to a subject or estimate their age (or value) of PBMCV, they can still be used within the overall data set to estimate curves for each gender. In most of the cases, these subjects do not have anything wrong with their data. Rather, they have a limited age range and have a similar pattern of bone growth throughout this range making it difficult to estimate an age and a value of peak bone growth. For more on this topic, refer to the Data Rules appendix graphs 4 through 6.

Histograms of the variables were examined (omitted here) to look at normality and shape of the distributions. Age of menarche was slightly left-skewed with a peak between 12 and 13. Age of PBMCV looked normal for females, but not for males. It looked closer to normal for the overall group than for males. For age of PHV, the overall graph looks normal, the male graph is close to normal, and the female graph is

left-skewed. Age of test is extremely right-skewed. The reason being is that they did a few year follow-up on the patients around 30 years of age. This is one of the main reasons that the data was limited to only include observations with age ≤ 22. All 3 height distributions are left-skewed, while all 3 weight distributions are right-skewed. Lastly, the TBBMC graphs do not look normal, particularly the male graph. The overall graph is a bit right-skewed.

In addition to univariate distributions, bivariate distributions were examined using scatterplots. Figure 3.1 has all subjects plotted. Black circles are for male subjects and red circles are for female subjects (this is standard throughout these graphs).Here is a graph of all subjects' age versus TBBMC. From the graph, the growth levels off around 18 or 20 years of age for most people. This is another reason to limit age to ≤ 22.

Figure 2.1.: Graph of Age vs TBBMC (in g) by Gender



In conclusion, there are several overarching characteristics of the Bailey data set. First, the females tend to reach the age of benchmarks (like PHV or PBMCV) before the males. Second, the males tend to be bigger than the females as a whole (e.g. with respect to height, weight, and TBBMC). Additionally, while the subjects had between 1 and 22 visits, they had TBBMC measured on only half of these. Therefore, for some individuals, a clear peak of bone growth was not evident (111 of the 251 subjects).

3. PROBABILITY MODEL FOR TOTAL BODY BONE MINERAL CONTENT (TBBMC)

This chapter is focused on finding a model to fit TBBMC. Initially individual data were examined to see what the relationship between age and TBBMC looked like. It has both upper and lower asymptotes, has non-negative values, is non-decreasing (over the desired age range), has a rate of growth that changed with age and was higher in the middle range of ages than toward the ends. Figure 3.2 shows individuals' TBBMC curves that exhibit the properties of upper and lower asymptotes, non-negative values, and a non-decreasing function. Figure 3.4 shows individuals that are examples of the sharper decrease in bone growth before and after peak. The properties of the growth curves were used to select an asymmetric logistic model. In addition to finding an appropriate model, this chapter examines the important values derived from the model including the age and values of Peak Bone Mineral Content Velocity (PBMCV) as well as their standard deviations.

Various functions with these properties were examined, which included the class of cumulative distribution functions (CDFs). The reason for that is CDFs are non-decreasing, non-negative, and they typically have a lower and upper asymptote as was displayed by the bone accumulation. Some of them also have a varying rate of growth. Some examples of potential CDFs to use are the logistic, the Gompertz, and the Weibull.

CDFs represent the cumulative probability for a distribution and are therefore always between 0 and 1 inclusive. TBBMC values in the data set went up to as high as about 3800 g or 3.8 kg. Therefore, the CDFs were generalized by adding an

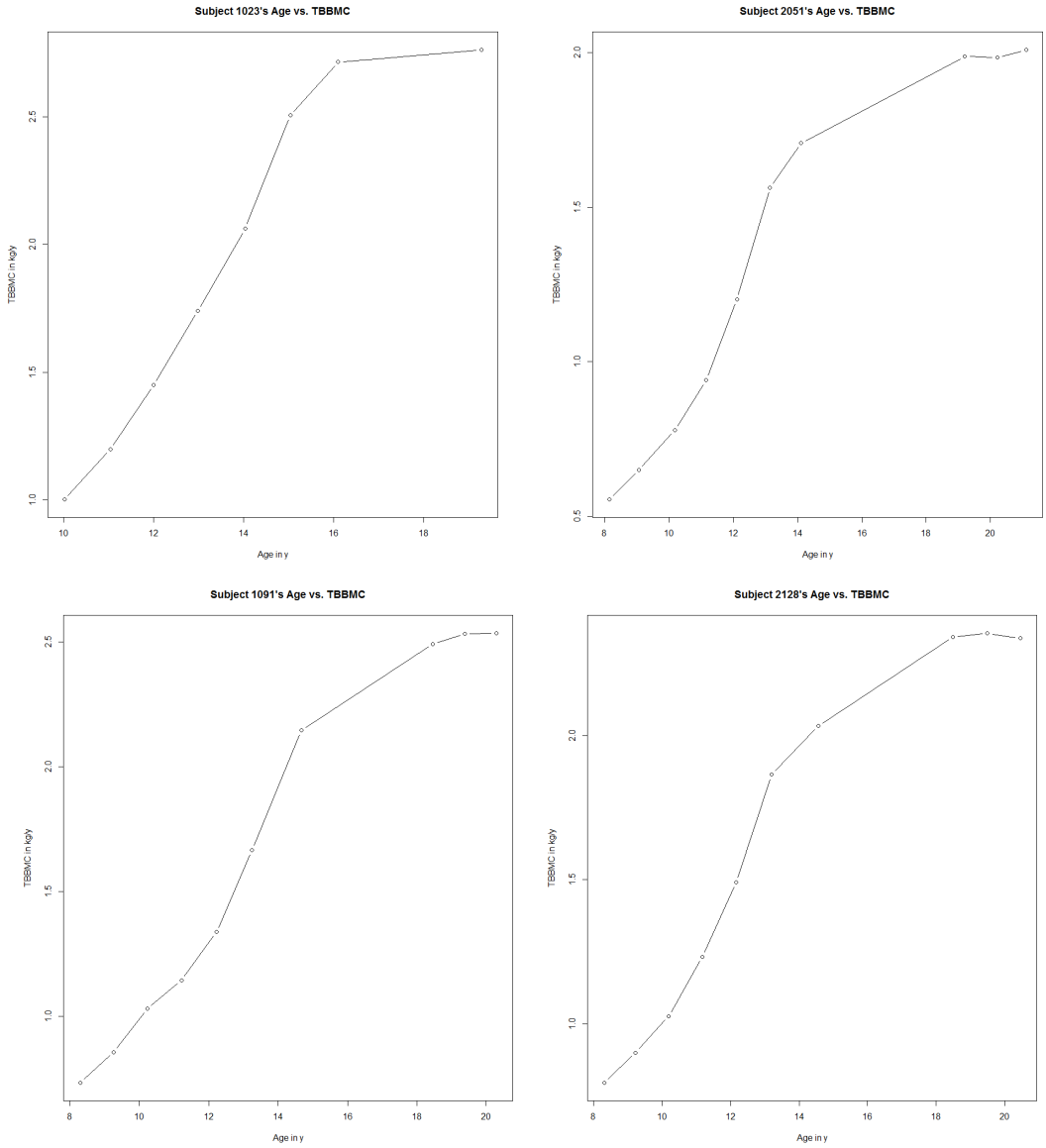


Figure 3.2.: Individual Subjects' TBBMC Plots

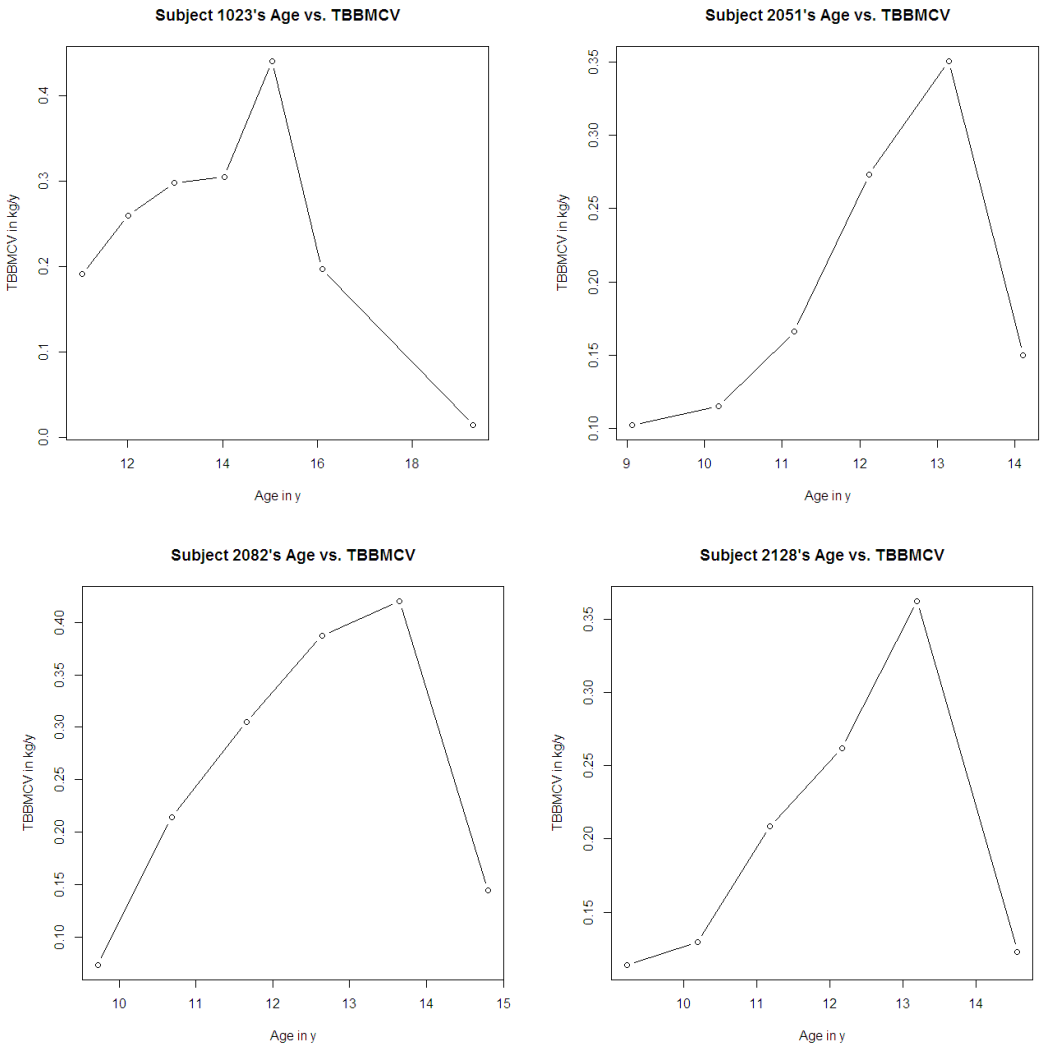


Figure 3.4.: Individual Subjects' TBBMCV Plots Exhibiting Sharp Decline After Their Peak

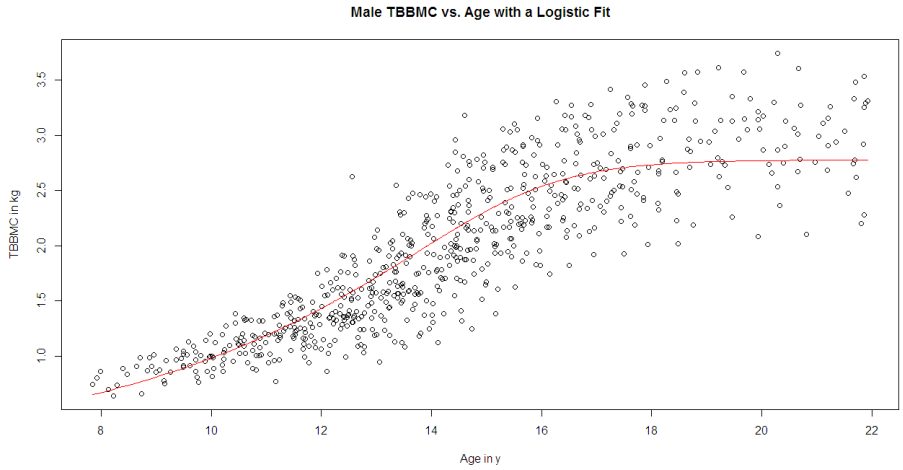


Figure 3.5.: Male TBBMC vs. Age with a Logistic Fit

Figure 3.6.: Male TBBMC vs. Age with a Gompertz Fit

upper asymptote parameter, β, so that now they go from 0 to β. The Figures 3.5 - 3.7 illustrate the male TBBMC data with fitted values from the aforementioned CDFs.

The age range of the data is about 8 to 22 years old. Since the subjects are not newborns (or close to it), the TBBMC values are not near 0. With this data set, they are from about .5 kg to about 3.8 kg. In Figures 2.3 - 2.5 they all fit the data well

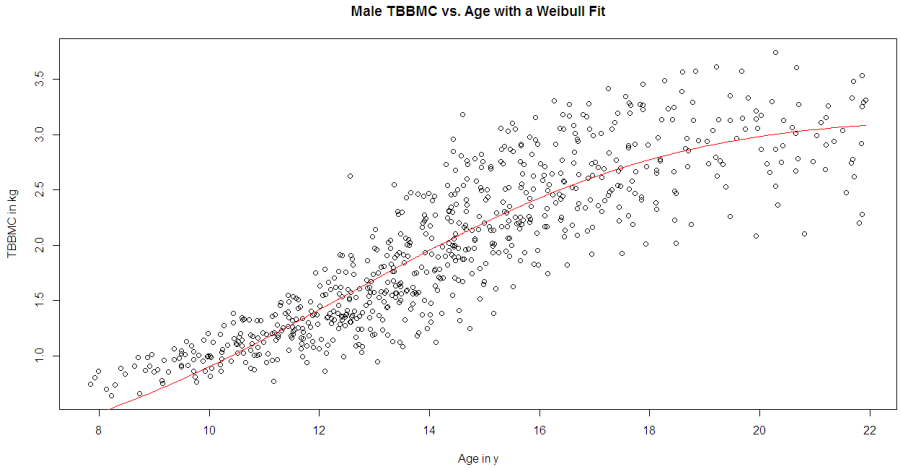


Figure 3.7.: Male TBBMC vs. Age with a Gompertz Fit

from about 10 to 20 years of age. The logistic fit does a better job of handling the lower ages (8-10) as the other fits drop off too quickly.

The purpose of modeling bone growth is to obtain 4 estimates. They revolve around Peak Bone Mineral Content Velocity (PBMCV), or the maximal rate of bone mass growth. The important values are: the age of PBMCV, the standard deviation (SD) of the age of PBMCV, the value of PBMCV, and the SD of PBMCV. The parameters were assumed to be jointly Normal with a mean and covariance structure to be estimated. This would allow the calculation of the SD of PBMCV (SD(PBMCV)). The goal is to use the distribution of PBMCV to calculate both an EAR and an RDA. Assuming a Normal Distribution, then the value of PBMCV would be both the mean and the median. Since PBMCV is the median, it is used to estimate the EAR of calcium, since the EAR is based off of a median. Additionally, PBMCV + 2 SD(PBMCV) is used to estimate the RDA of calcium, since the RDA is based off the $97^{th} - 98^{th}$ percentile.

3.1 Logistic Function for Bone Accumulation

After investigation of the residuals as well as individual curves of TBBMC versus age, it was verified that the rate of growth decreased faster after the age of PBMCV than it increased beforehand. Therefore, a use of a symmetric logistic model was deemed inappropriate. To incorporate asymmetry, the parameter γ was introduced. (For this version to display the requisite growth behavior, γ would need to be between 0 and 1.) It was found that γ being a random variable or a constant provided similar results. Additionally, removing γ from the multivariate normality assumption improved the normality of important output variables like age of PBMCV. For this reason, it was set as a constant. The value of γ is .2, which was calculated as the average of the male and female values (since they were not shown to be different).

While forcing γ to be a constant did not affect the gender stratified model, it had an impact on the individual subjects' parameterizations. However, the goal is to generalize by age(s) not subject, so this change is not troublesome.

The set of parameters, Θ, for the asymmetric logistic is $\Theta = (\mu, \beta, \sigma^{-1})$ with $\gamma = .2$. In the symmetric model, μ represented the age of maximum growth. In the asymmetric model, it plus a value is the age of maximum growth, with μ being the majority of the age of PBMCV. The parameter β represents the upper asymptote, which is the amount of TBBMC after growth is complete. Lastly, σ^{-1} is a variability measure.

This model has some nice properties. To begin with, the assumption of joint normality of the parameters was not rejected. Similarly, the values of PBMCV, the age of PBMCV, and the average TBBMCV (by age group) did not reject the normality assumption (by gender). Figures 3.9 and 3.11 show qqplots of these last 4 variables. Figure 3.9 represents the males' variables while Figure 3.11 represents the females' variables.

Normality of the important values as well as the simplification in the calculation of SD(PBMCV) lead to the use of this as the final model. The final model has bone accumulation (TBBMC) as Y and age as X. For an individual,

$$Y = F(x) = \beta(1 + exp(-\sigma^{-1}(x - \mu)))^{-.2},$$

and rate of bone growth (TBBMCV) as

$$TBBMCV = f(x) = \frac{d}{dx}F(x) = .2\sigma^{-1}\beta\frac{exp(-\sigma^{-1}(x - \mu))}{(1 + exp(-\sigma^{-1}(x - \mu)))^{1.2}}.$$

Let $\Theta = (\mu, \sigma^{-1}, \beta)$ (with $\gamma = .2$) be the parameters' population mean vector of F(x).

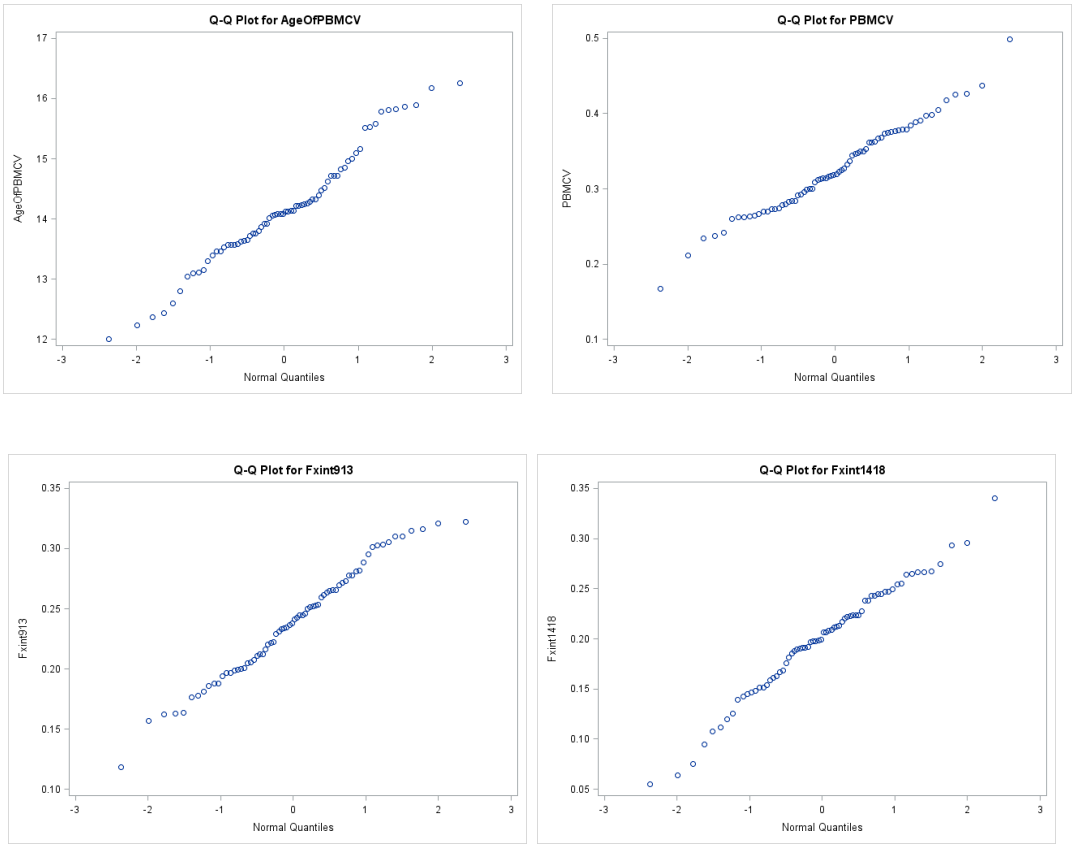


Figure 3.9.: Male QQ Plots

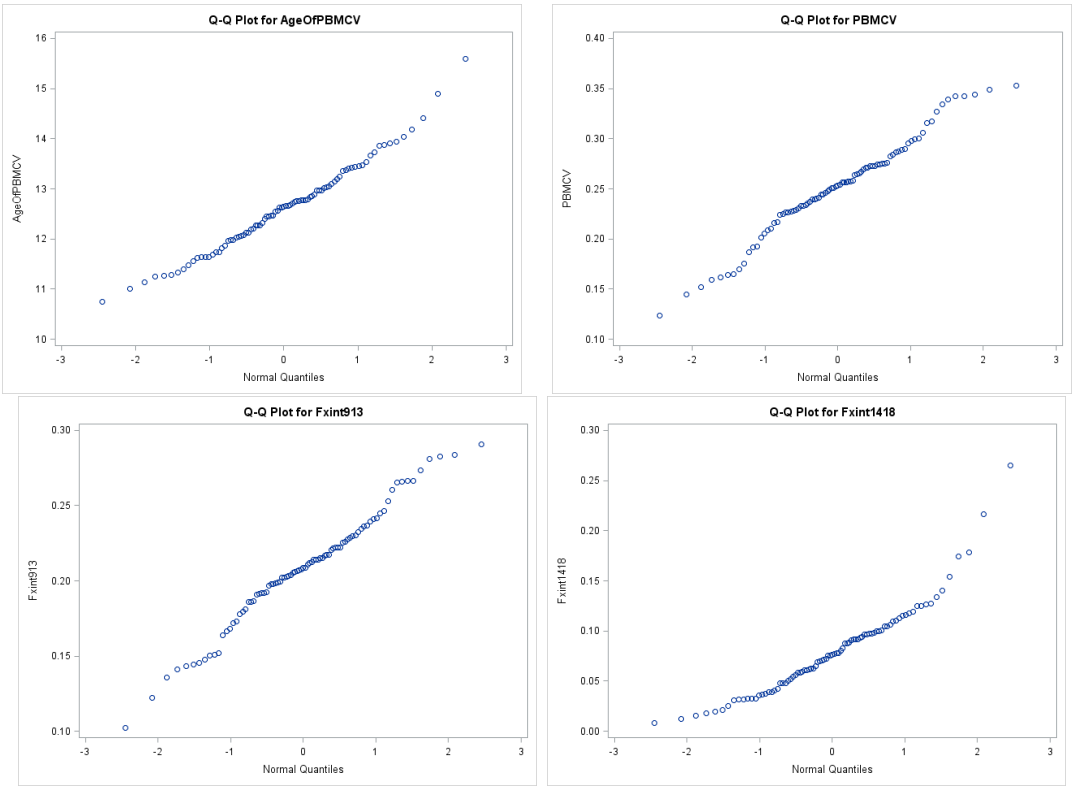


Figure 3.11.: Female QQ Plots

The model (by gender) is TBBMC $\sim$ N(F(x), ϵ^2) and $\widehat{\Theta} \sim$ N(Θ, Σ) where ϵ^2 is a 1-dimensional error term, $\widehat{\Theta}$ is the estimated parameter mean vector, and Σ is the parameter covariance matrix (both 3 by 3).

With the asymmetric model, the next section will explore the various values of interest. Namely, they will study the age of peak bone growth (PBMCV), the value at this age, and the variance of this value.

3.2 Important Values

3.2.1 Age of PBMCV

An important question is what would be the estimate for the age of the maximum growth rate, i.e. the x value at which its derivative is maximized. Specifically, what is the age at which peak Total Body Bone Mineral Content Velocity (TBBMCV) occurs?

As above, let

$$F(x) = \beta(1 + exp(-\sigma^{-1}(x - \mu)))^{-.2}.$$

Then,

$$\frac{d}{dx}F(x) = f(x) = .2\sigma^{-1}\beta\frac{exp(-\sigma^{-1}(x-\mu))}{(1+exp(-\sigma^{-1}(x-\mu)))^{1.2}}.$$

Since the goal is to find the x value which maximizes this, the next step is to take the derivative of this and set it equal to 0. For ease of notation, let g(x) = $exp(-\sigma^{-1}(x - \mu))$.

Here, the quotient rule of derivatives was used to get

$$f'(x) = .2\sigma^{-1}\beta\frac{\sigma^{-1}g(x)(1+g(x))^{1.2} - g(x)(1.2)(1+g(x))^{.2}(-\sigma^{-1}g(x))}{(1+g(x))^{2.4}}.$$

This can be reduced to

$$f'(x) = .2\sigma^{-1}\beta(-\sigma^{-1}g(x))(\frac{1}{(1+g(x))^{1.2}} - \frac{g(x)(1.2)}{(1+g(x))^{2.2}}).$$

Setting this equal to 0 and solving for x, the first part goes away and what remains is:

$$\frac{1}{(1+g(x))^{1.2}} = \frac{g(x)(1.2)}{(1+g(x))^{2.2}}.$$

Multiplying by $(1+g(x))^{2.2}$ yields

$$1 + g(x) = g(x)(1.2)$$

which is $1 = g(x).2$ or $\frac{1}{.2} = g(x)$. Replacing g(x) with $exp(-\sigma^{-1}(x-\mu))$, results in

$$ln(\frac{1}{.2}) = -\sigma^{-1}(x-\mu)$$

or

$$x = \mu - \frac{ln(1/.2)}{\sigma^{-1}}.$$

But, by one of the properties of logarithms (namely $log(x^n) = nlog(x)$), this also becomes

$$x = \mu + \frac{ln(.2)}{\sigma^{-1}}.$$

For the asymmetric model used herein,

$$Age \quad PBMCV = \mu + \frac{ln(.2)}{\sigma^{-1}}.$$

3.2.2 Value of PBMCV

For this part, the fact that g(x) = $\frac{1}{.2}$ at x = age of PBMCV is used. Since,

$$f(x) = .2\sigma^{-1}\beta\frac{g(x)}{(1+g(x))^{1.2}},$$

knowing that g(x) = $\frac{1}{.2}$, this becomes

$$f(\mu + \frac{ln(.2)}{\sigma^{-1}}) = .2\sigma^{-1}\beta\frac{1}{.2}(1+\frac{1}{.2})^{-1.2}.$$

Simplifying, the value of PBMCV is

$$\beta\sigma^{-1}(6)^{-1.2}.$$

For the asymmetric model used herein,

$$PBMCV = 6^{-1.2}\beta\sigma^{-1}.$$

3.2.3 Delta Method for Variances

The values just computed were the age of PBMCV and the value of PBMCV. The variances, or standard deviations, of these values are also desired. First is the calculation of Var(PBMCV). As a function of the parameters,

$$PBMCV = 6^{-1.2}\beta\sigma^{-1}.$$

Looking at the k-multidimensional Delta Method in Casella and Berger [6],

$$var(g(T)) \approx \sum_{i=1}^{k}[g_i'(\theta)]^2 Var_\theta(T_i) + 2\sum_{i>j} g_i'(\theta)g_j'(\theta)Cov_\theta(T_i, T_j).$$

Here $\Theta = (\mu, \sigma^{-1}, \beta)$ and g(T) = $\beta\sigma^{-1}$. The partial derivative with respect to β is σ^{-1} and the partial derivative with respect to σ^{-1} is β. This leads to:

$$var(\beta\sigma^{-1}) \approx \sigma^{-2}Var(\beta) + \beta^2 Var(\sigma^{-1}) + 2\beta\sigma^{-1}Cov(\beta, \sigma^{-1}).$$

It is known that Var(aX) = a^2 Var(X) when a is a constant. Since γ is treated as a constant in the final model, then

$$\begin{aligned} Var(PBMCV) &= 6^{-2.4}Var(\beta\sigma^{-1}) \\ &= 6^{-2.4}[(\sigma^{-1})^2 Var(\beta) + \beta^2 Var(\sigma^{-1}) + 2\beta\sigma^{-1}Cov(\beta, \sigma^{-1})]. \end{aligned}$$

Next is the calculation of Var(Age of PBMCV). As a function of the parameters,

$$AgeofPBMCV = \mu + ln(.2)\left(\sigma^{-1}\right)^{-1}.$$

The partial derivative with respect to μ is 1 and the partial derivative with respect to σ^{-1} is $-ln(.2)/\sigma^{-2}$. Using these within the Delta Method formula results in

$$Var(AgeofPBMCV) = Var(\mu) + \frac{ln(.2)^2}{\sigma^{-4}}Var(\sigma^{-1}) - 2\frac{ln(.2)}{\sigma^{-2}}Cov(\mu, \sigma^{-1}).$$

3.3 SAS Implementation and Model Estimates

Let Θ be the set of parameters. It should be noted that for each parameterization, 3 models were run relating to all subjects get the same Θ, Θ varies by gender, and

Θ varies by person. While it is computationally expensive to have Θ vary by person, and this would not generalize to other data sets, this is used to check the normality assumptions and for simulation purposes. Additionally, the individual subjects' models averages and variances were used as starting values within the other 2 models.

As previously stated, one restriction made on the data set was that only ages ≤ 22 were used. This took the data down from 2,106 observations to 1,518. Since the model has 3 parameters, it was first ran on individuals with 5 or more observations to eliminate overfitting. Under this restriction (as well as the age restriction), there are 189 subjects and 1,356 observations. In SAS, the model (by gender) is TBBMC $\sim$ N(F(x), ϵ^2) and $\widehat{\Theta} \sim$ N(Θ, Σ) where F(x) is the asymmetric logistic $(\beta(1 + exp(-\sigma^{-1}(x - \mu)))^{-.2})$, ϵ^2 is a 1-dimensional error term, $\widehat{\Theta}$ is the estimated parameter mean vector, and Σ is the parameter covariance matrix (both 3 by 3).

3.3.1 Model Estimates and TBBMC(V) Curves

Running the asymmetric model with $\gamma = .2$ yielded the following:

$$\widehat{TBBMC}(male) = 2.78 * (1 + exp(-.96(Age - 15.41)))^{-.2}$$

$$\widehat{TBBMC}(female) = 2.14 * (1 + exp(-.96(Age - 14.03)))^{-.2}$$

Or, you could think of the parameterization as:

$$\Theta = (\mu, \beta, \sigma^{-1}, \gamma).$$

$$\widehat{\Theta}_m = (15.41, 2.78, .96, .2).$$

$$\widehat{\Theta}_f = (14.03, 2.14, .96, .2).$$

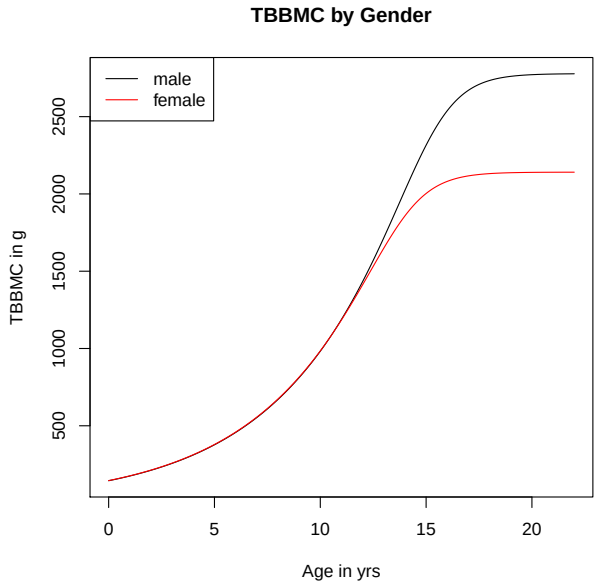
Additionally, they have the following estimated covariance structures:

These distributions yield the following F(x) and f(x) curves:

	μ_f	β_f	σ_f^{-1}
μ_f	0.8732	0.0167	-0.0808
β_f	0.0167	0.1062	-0.0081
σ_f^{-1}	-0.0808	-0.0081	0.0186

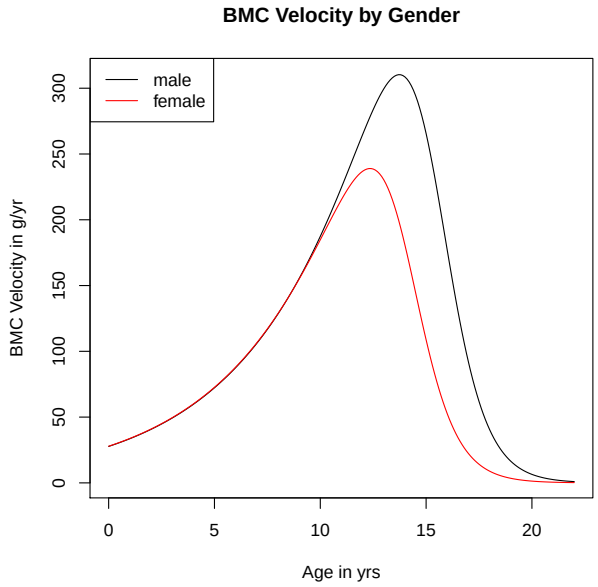
	μ_m	β_m	σ_m^{-1}
μ_m	2.0119	0.2208	-0.2616
β_m	0.2208	0.2117	-0.0330
σ_m^{-1}	-0.2616	-0.0330	0.0522

Figure 3.12.: F(x), or TBBMC, with the Asymmetric Model



Figures 3.12 and 3.13 highlight some of the differences across genders. Namely, the upper asymptote (β) is higher for males, so their right side of TBBMC is higher. Also notice that the females peak earlier (best seen in the TBBMCV graph), and that their peak is smaller than that of the males.

Figure 3.13.: f(x), or TBBMCV, with the Asymmetric Model



3.3.2 Important Values Estimated

An important value is PBMCV. Recall,

$$PBMCV = 6^{-1.2}\beta\sigma^{-1},$$

so it was .31032 and .23898 (kg/y) for males and females respectively. To illustrate this calculation, below is the males' estimate:

$$.31032 = 6^{-1.2} * 2.7783 * .959.$$

This is simply the above formula with $\beta = 2.7783$ and $\sigma^{-1} = .959$. The reason for the measurements are that β is measured in kg and σ^{-1} in y^{-1}.

Also of interest is the standard deviation of PBMCV. Recall,

$$Var(PBMCV) = 6^{-2.4}[(\sigma^{-1})^2 Var(\beta) + \beta^2 Var(\sigma^{-1}) + 2\beta\sigma^{-1} Cov(\beta, \sigma^{-1})].$$

Again, as an example, below is the illustration of Var(PBMCV) for the males (in $(kg/y)^2$). Therefore, the males' Var(PBMCV) is

$$.0057 = 6^{-2.4}[.959^2 * .2117 + 2.7783^2 * .0522 + 2 * 2.7783 * .959 * -.0330].$$

Next, the age of PBMCV as well as its standard deviation are also important. Recall,

$$AgeofPBMCV = \mu + \frac{ln(.2)}{\sigma^{-1}}.$$

The values are 13.74 and 12.35 years old for males and females respectively. The calculation for males is

$$AgeofPBMCV = 13.74 = 15.4141 + \frac{ln(.2)}{.959}.$$

As for the variance of PBMCV, the formula is

$$Var(PBMCV) \approx Var(\mu) + \left(\frac{ln(.2)}{\sigma^{-2}}\right)^2 - 2\frac{ln(.2)}{\sigma^{-2}}Cov(\mu, \sigma^{-1}).$$

The standard deviations of the age of PBMCV are .65 and 1.26 years respectively for females and males. As an example, this calculation for males is

$$SD(AgeofPBMCV) = 1.26 = \sqrt{2.0119 + \left(\frac{ln(.2)}{.959^2}\right)^2 - 2\frac{ln(.2)}{.959^2} * -.2616}.$$

Table 1 in Bailey et al.'s 2000 paper can be used for comparison purposes. There they give the values of PBMCV and its SD in parentheses for males and females as 407 (93) and 322 (66) g/y respectively. [7] We have 310 (76) and 239 (45). Both of these estimates are lower and have lower standard deviations than those by Bailey et al. Also, they give the values of the age of PBMCV and its SD in parenthesis. Their values are 14 (1) and 12.5 (.9) for males and females respectively. We have 13.74 (1.26) and 12.35 (.65) for males and females respectively. These values are very similar.

Lastly, here is a table summarizing the important values by gender.

Table 3.1.: Important Values Estimates by Gender

	Males	Females
PBMCV (in g/y)	310.33	238.98
SD(PBMCV) (in g/y)	75.65	45.02
Age of PBMCV (in y)	13.74	12.35
SD(Age of PBMCV) (in y)	1.26	.65

3.3.3 Delta Method Check via Simulation

This section has further investigation of simulated results versus the theoretical Delta Method for the variance of PBMCV or $\text{Var}(\beta\sigma^{-1}(6)^{-1.2}$. For this simulation, for each gender 10,000,000 values of Θ was simulated using the multivariate normal structure of Θ. The inputs were Θ_m, Θ_f, Σ_m, and Σ_f. At the end, means and covariances of the parameters were compared to the inputs to check the simulation. Since these were accurate to 2 or 3 decimals, age and values of PBMCV for each simulated Θ. These vectors of 10 million observations were then used for 2 values, their mean and variance.

There was some debate whether to use plug-in estimates or expected value estimates within the Delta Method calculation. Referring to examples 3.4 and 3.5 in ATSP [8], in place of X^2 he uses μ^2 or the estimate of X squared. He does not use an expectation style estimate (which would include the variance of X too). Therefore, plug-in estimates will be used throughout the Delta Method material. This means that for something like σ^{-2}, $(\hat{\sigma^{-1}})^2$ will be used and not an expectation (which would be this estimate plus the $\text{Var}(\sigma^{-1})$.

Table 3.2.: Delta Method Check Via Simulation

		plug-in	simulation
Males	Age PBMCV (in y)	13.74	13.73
	Value PBMCV (in kg/y)	0.3103	0.3103
	Var(PBMCV) (in $(\text{kg/y})^2$)	0.0057	0.0059
Girls	Age PBMCV (in y)	12.35	12.35
	Value PBMCV (in kg/y)	0.239	0.239
	Var(PBMCV) (in $(\text{kg/y})^2$)	0.0020	0.0021

These estimates were the same to 3 decimals.

3.3.4 Assumption Checking

With the symmetric models, multivariate normality (MVN) was lacking. Mardia's skewness and kurtosis as well as Heinze-Zeinkler tests of MVN were run for each model. Only the asymmetric model did not reject the MVN structure of the parameters. This is one of the reasons this is the final model.

More specifically, all of the male tests did not reject multivariate normality. The female kurtosis test did not reject multivariate normality, neither did the Henze-Zirkler test. However, the Mardia (small-sample) skewness did reject normality.

In conclusion, the asymmetric logistic function employed above is 1 of several attempted. It was chosen for 3 reasons. It allowed the use of a random statement to estimate the covariance structure of the parameters. This is important for converting from EAR to RDA. Furthermore, the parameterization was shown to be about multivariate normal. One or both of these statements were missing for the other models attempted. Lastly, the ages and values of PBMCV (for the individual subjects) as well as their standard deviations did not reject the normality assumption. Again, this is important for estimating RDA. Lastly, this model was able to incorporate individuals who did not exhibit clear peaks in bone growth.

3.3.5 Average Rate of Bone Growth By Age Group and Gender

The current EAR and RDA values for calcium intake were formed by averaging over the age groups (9-14 and 14-18). This section will examine average bone growth by these age groups and gender. For the model, rate of bone growth, or TBBMCV is estimated by f(x). To calculate the average of TBBMCV over an interval (say from a to b), the formula would be

$$\int_a^b \frac{1}{b-a} f(x)dx.$$

Since the interest lies in specific age ranges, a and b would be considered constants. Therefore, the average of TBBMCV over the ages a to b is

$$\frac{1}{b-a}\int_a^b f(x)dx = \frac{1}{b-a}[F(b) - F(a)].$$

Another way to look at this, is that F(x) represents TBBMC, or the total mass of bone in the body at age x. So, F(b) - F(a) is the total change in bone mass from age a to age b. Dividing by b - a, the length of time considered, changes it from total change in bone mass to average change in bone mass. So, there are 2 primary values of interest for each gender, $\frac{1}{14-9}[F(14) - F(9)]$ and $\frac{1}{18-14}[F(18) - F(14)]$. The values of these can be calculated directly from model. In turn, the values can be used to estimate an EAR for calcium intake. However, a standard deviation of these values is necessary to computer an RDA for calcium intake.

The following is material related to the Delta Method for the difference in F(x) values. In this case, both $F_1(\Theta)$ and $F_2(\Theta)$ will be the same with respect to Θ. The only difference will be the x value input. Both functions are

$$F(\Theta) = \beta(1 + exp(-\sigma^{-1}(x - \mu)))^{-\gamma}.$$

Let

$$\Theta = \left(\mu, \beta, \sigma^{-1}\right).$$

For ease of reading, let

$$g(x) = exp(-\sigma^{-1}(x - \mu)).$$

The derivative vector is

$$\frac{dF(\Theta)}{d\Theta} = \left(-\gamma\beta\sigma^{-1}g(x)(1 + g(x))^{-\gamma-1}, (1 + g(x))^{-\gamma}, \left(\gamma\beta(x - \mu)g(x)(1 + g(x))^{-\gamma-1}\right)\right).$$

The value of Var$(F(\Theta))$ can be found with matrix algebra (using the first order Delta method); namely,

$$Var(F(\Theta)) \approx \left(\frac{dF(\Theta)}{d\Theta}\right)\Sigma\left(\frac{dF(\Theta)}{d\Theta}\right)'.$$

As an example, $\hat{\Theta}_{males} = (15.4141, 2.7783, .959)$ with $\gamma = .2$. For 9 year old males, the derivative vector is:

$$\left(\frac{dF(\Theta)}{d\Theta}\right) = (-.1553, .2921, -1.0389).$$

Additionally,

$$\hat{\Sigma}_{males} = \begin{bmatrix} 2.0119 & .2208 & -.2616 \\ .2208 & .2117 & -.0330 \\ -.2616 & -.0330 & .0522 \end{bmatrix}$$

Putting this altogether, the first order approximation of Var(F(Θ)) when x = 9 is

$$(-.1553, .2921, -1.0389) \begin{bmatrix} 2.0119 & .2208 & -.2616 \\ .2208 & .2117 & -.0330 \\ -.2616 & -.0330 & .0522 \end{bmatrix} (-.1553, .2921, -1.0389)'$$

So,

$$Var(F(\Theta)) \approx .0385$$

at x = 9.

A simulation was run with a sample size of 10,000,000 drawn from N($\hat{\Theta}_{males}$,$\hat{\Sigma}_{males}$). For each of these values, F(9), F(14), and F(18) were all calculated. Additionally, variances and covariances of these 3 values were also computed. As an example,

$$Var(F(9)) = .0530$$

in the simulation. This is about 37.6% higher than the value computed from the Δ method. The discrepancy is due to the fact that $F(\Theta)$ is non-linear in both μ and σ^{-1} and has many derivatives with respect to these parameters, not just the first derivative. Therefore, it is appropriate to add higher order terms. A formula to include the first k orders would be:

$$Var(F(\Theta)) = \sum_{i=1}^{k} \left(\frac{d^i F(\Theta)}{d^i \Theta * i!}\right) \Sigma \left(\frac{d^i F(\Theta)}{d^i \Theta * i!}\right)'.$$

Higher orders means using higher orders of derivatives. It also means including factorials. This is based off of Taylor Series expansion of functions.

As an example of a higher order term,

$$\left(\frac{d^2F(\Theta)}{d^2\mu}\right) = -\gamma\beta\sigma^{-1}[\sigma^{-1}g(x)(1+g(x))^{-\gamma-1} + g(x)^2(1+g(x))^{-\gamma-2} * (-\gamma-1) * \sigma^{-1}]$$

$$\left(\frac{d^2F(\Theta)}{d^2\beta}\right) = 0$$

$$\left(\frac{d^2F(\Theta)}{d^2\sigma^{-1}}\right) = \gamma\beta(x-\mu)[g(x)(1+g(x))^{-\gamma-1} * (\mu - x)$$
$$+ g(x)^2(-\gamma-1)(1+g(x))^{-\gamma-2}(\mu-x)]$$

For the males at age 9,

$$\left(\frac{d^2F(\Theta)}{d^2\Theta}\right) = (.0294, 0, 1.3157).$$

The resulting matrix multiplication (including the factorials) results in a value of .0180. So,

$$Var(F(\Theta)) \approx .0385 + .0180 = .0565$$

at x = 9. This is closer to the simulated value (about 6% different as opposed to about 38%). The values from the (2^{nd} order) Δ method for Var(F(14)) and Var(F(18)) were .1639 and .1868 respectively.

While the aforementioned variances are a part of the work in this section, the goal is to estimate the variance of the difference in F(Θ) values. Specifically, the goal is to calculate:

$$Var(F(\Theta)|_{x_1} - F(\Theta)|_{x_2}) = Var(F(\Theta)|_{x_1}) + Var(F(\Theta)|_{x_2}) - 2Cov(F(\Theta)|_{x_1}, F(\Theta)|_{x_2}).$$

The covariance of 2 functions (with the same Θ) via the Delta Method is needed. The formula is:

$$Cov(F_1(\Theta), F_2(\Theta)) = \sum_i \left(\frac{dF_1(\Theta)}{d\Theta_i}\right)\left(\frac{dF_2(\Theta)}{d\Theta_i}\right) Var(\Theta_i) + \sum_i \sum_{j \neq i} \left(\frac{dF_1(\Theta)}{d\Theta_i}\right)\left(\frac{dF_2(\Theta)}{d\Theta_j}\right) Cov(\Theta_i, \Theta_j). \tag{3.1}$$

This formula is more general than what is needed here, since the F(Θ) are the same, they are just evaluated at different x values. This will make the computations slightly easier. The derivatives would be the same as what was calculated above. The mathematics are omitted here. The following results all reflect work done using the Δ Method. For the covariance of F(Θ) for males aged 9-14 and for males aged 14-18, the values were .0488 and .1169. Similarly, the covariance of F(Θ) for females aged 9-14 and for females aged 14-18, were .0507 and .0903.

Revisiting, the formula

$$Var(F(\Theta)|_{x_1} - F(\Theta)|_{x_2}) = Var(F(\Theta)|_{x_1}) + Var(F(\Theta)|_{x_2}) - 2Cov(F(\Theta)|_{x_1}, F(\Theta)|_{x_2}),$$

results in the following table.

Table 3.3.: Variances of the Difference in F(Θ) Values by Age and Gender

	Var(F(14) - F(9))	Var(F(18) - F(14))
Males	0.1228	0.0363
Females	0.0966	0.0282

The following tables show the values of F(Θ) as well as the values necessary to compute the variances of the average difference in F(Θ) values based on the age group and gender.

The last 2 columns of the previous 4 tables are summarized in this table.

Table 3.4.: Average TBBMCV for 9-14 Year Old Males

unit	kg	kg	kg	kg/y
	F(9)	F(14)	F(14) - F(9)	$\frac{F(14)-F(9)}{14-9}$
value	0.81	2.02	1.21	0.2424
var	0.06	0.16	0.12	0.0049
cov or sd	0.05	0.05	0.35	0.0701

Table 3.5.: Average TBBMCV for 14-18 Year Old Males

unit	kg	kg	kg	kg/y
	F(14)	F(18)	F(18) - F(14)	$\frac{F(18)-F(14)}{18-14}$
value	2.02	2.73	0.71	0.1776
var	0.15	0.18	0.10	0.0060
cov or sd	0.12	0.12	0.31	0.0777

Table 3.6.: Average TBBMCV for 9-14 Year Old Females

unit	kg	kg	kg	kg/y
	F(9)	F(14)	F(14) - F(9)	$\frac{F(14)-F(9)}{14-9}$
value	0.81	1.86	1.04	0.2087
var	0.03	0.10	0.04	0.0015
cov or sd	0.05	0.05	0.19	0.0381

Table 3.7.: Average TBBMCV for 14-18 Year Old Females

unit	kg	kg	kg	kg/y
	F(14)	F(18)	F(18) - F(14)	$\frac{F(18)-F(14)}{18-14}$
value	1.86	2.13	0.27	0.0684
var	0.10	0.10	0.03	0.0018
cov or sd	0.09	0.09	0.17	0.0420

Table 3.8.: [Average] Differences in F(Θ) by Age and Gender in kg or kg/y

	F(14) - F(9)	$\frac{F(14)-F(9)}{14-9}$	F(18) - F(14)	$\frac{F(18)-F(14)}{18-14}$
Male's Value	1.21	0.2424	0.71	0.1776
Male's Variance	0.12	0.0049	0.10	0.0060
Female's Value	1.04	0.2087	0.27	0.0684
Female's Variance	0.04	0.0015	0.03	0.0018

The following table converts the average and standard deviations of the differences in F(Θ) to Ca mg/d, which will be used directly in the formulas for converting to intake needed.

Table 3.9.: Average and SD of the Differences in F(Θ) in Ca mg/d

	$\frac{F(14)-F(9)}{14-9}$	$SD\left(\frac{F(14)-F(9)}{14-9}\right)$	$\frac{F(18)-F(14)}{18-14}$	$SD\left(\frac{F(18)-F(14)}{18-14}\right)$
Males	213.81	61.83	156.72	68.55
Females	184.08	33.61	60.30	37.05

4. RELATION TO CALCIUM REQUIREMENTS

The main purpose of modeling bone accrual, or total body bone mineral content (TBBMC), is to be able to find a distribution for the growth and have an estimate of the value of maximum growth and its standard deviation. Peak bone mineral content velocity, or PBMCV, is the value of maximum growth. Having both the mean and standard deviation of PBMCV allows the calculation of both an EAR and an RDA for calcium. This chapter focuses on the values of PBMCV as well as average values of TBBMC velocity to convert them to EAR and RDA by various methods.

For the model, γ was fixed at .2, $\widehat{\Theta}_m = (15.4141, 2.7783, .959)$, and $\widehat{\Theta}_f = (14.0317, 2.1409, .9584)$, where $\Theta = (\mu, \beta, \sigma^{-1})$. Additionally, the covariance structures are:

$$\widehat{\Sigma}_{female} = \begin{array}{c} \\ \mu_f \\ \beta_f \\ \sigma_f^{-1} \end{array} \begin{array}{c} \begin{array}{ccc} \mu_f & \beta_f & \sigma_f^{-1} \end{array} \\ \begin{pmatrix} 0.8732 & 0.01668 & -0.08076 \\ 0.01668 & 0.1062 & -0.00813 \\ -0.08076 & -0.00813 & 0.01859 \end{pmatrix} \end{array}$$

$$\widehat{\Sigma}_{male} = \begin{array}{c} \\ \mu_m \\ \beta_m \\ \sigma_m^{-1} \end{array} \begin{array}{c} \begin{array}{ccc} \mu_m & \beta_m & \sigma_m^{-1} \end{array} \\ \begin{pmatrix} 2.0119 & 0.2208 & -0.2616 \\ 0.2208 & 0.2117 & -0.03296 \\ -0.2616 & -0.03296 & 0.05219 \end{pmatrix} \end{array}$$

Using this model, recall the values of interest are

$$PBMCV = 6^{-1.2}\beta\sigma^{-1},$$

and

$$Var(PBMCV) = 6^{-2.4}[(\sigma^{-1})^2 Var(\beta) + \beta^2 Var(\sigma^{-1}) + 2\beta\sigma^{-1} Cov(\beta, \sigma^{-1})].$$

For males,

$$PBMCV = 6^{-1.2} * 2.7783 * .959 = .31033$$

and

$$Var(PBMCV) = 6^{-2.4}[.959^2 * .2117 + 2.7783^2 * .05219 + 2 * 2.7783 * .959 * -.03296] = .0057.$$

These values are measured in kg/y and $(\text{kg/y})^2$ respectively. Therefore,

$$Males' \quad PBMCV(in \quad kg/y) \sim N(\mu = .31033, \sigma = \sqrt{.0057} = .07565).$$

Similarly,

$$Females' \quad PBMCV(in \quad kg/y) \sim N(\mu = .23898, \sigma = \sqrt{.0020} = .04502).$$

What is needed are these distributions converted to mg/d and compensating for the amount of bone that is calcium. Let G be a random variable describing the amount of calcium (in mg/d) needed for the growth of bone at the peak rate in a population of interest (e.g. females or males). Since this is just a constant times a Normal distribution, it too will be Normal. The constant will be $\frac{1000000}{365} * .322$. The fraction converts from kg/y to mg/d. The .322 represents the portion of bone mass that is attributed to calcium. [1] Now,

$$G_{males} \sim N(\mu = 273.77, \sigma = 66.74),$$

and

$$G_{females} \sim N(\mu = 210.83, \sigma = 39.72).$$

Let g_p be the p-th percentile of the distribution of G. Let s be the amount of calcium (in mg/d) lost by sweat and let b be the balance (in mg/d) corresponding to a given calcium intake and particular subject characteristics. Then,

$$b - s = g_p$$

and this relationship can be used to determine the intake required for the specified percentile, p. Of particular interest are the intakes corresponding to p=.5, the EAR,

and p=.975, the RDA. The μ of G is $g_{.5}$ and will be the estimates for the retention at the EAR level of calcium. For a Normal Distribution, the 97.5^{th} percentile is about 2 standard deviations above the mean, or $\mu + 2\sigma$, and this will be used for $g_{.975}$. Doing this for G for both the males and females' respectively, yields

$$273.77 + 2 * 66.74 = 407.25$$

and

$$210.83 + 2 * 39.72 = 290.27.$$

The values of 407.25 and 290.27 will be the estimates for the retention at the RDA level of calcium. Table 4.1 is a summary of this information.

Table 4.1.: Important Calcium Retention Values from Our Model

	Males	Females
$g_{.5}$ in mg/d	273.77	210.83
SD(G) in mg/d	66.74	39.72
$g_{.975}$ in mg/d	407.25	290.27

These values will be a constant throughout this section as they are from the model. They will be used in various formulas in established papers to relate them back to the appropriate intake level. Some of these formulas incorporate race, age, gender, and other explanatory variables as this has been shown to affect the intake and retention relationship. While the data used to establish PBMCV is only for white Canadian children, it will be used as a basis for these models. Therefore, one assumption is that the maximum growth rates' distributions are constant across these variables. Future work will test this assumption.

In addition to the comparison to the current DRI values, there is a comparison to the values Bailey et al. They had values of PBMCV as 407 and 322 g/y for males

and females respectively, with standard deviations of 93 and 66 (g/y). Converting these to mg/d of calcium (as before), we get the following table:

Table 4.2.: Important Calcium Retention Values from Bailey et al.

	Males	Females
$g_{.5}$ in mg/d	359.05	284.07
SD(G) mg/d	82.04	58.22
$g_{.975}$ in mg/d	523.14	400.52

4.1 Calcium Retention to Intake Formulas Using PBMCV

In these formulas, the word calcium will be omitted from retention and intake to make them a bit easier to read. In all of these models, intake is considered the explanatory variable and retention the response. However, all of the models will be inverted because the growth curves predict retention and that will be used to find intake levels. Intake and retention both will be measured in mg/d.

4.1.1 Factorial Method

The first formula used is the one listed in the calcium requirements by the Institute of Medicine (IOM) in their Dietary Reference Intakes guide for calcium. In general, the factorial method has

$$retention = \%absorption * intake - losses.$$

Inverting this formula, the factorial method formula for relating the calcium retained to intake is

$$intake = \frac{retention + losses}{\%absorption}. \tag{4.1}$$

The current values of losses of 290 and 273 mg/d for males and females respectively and a 38% absorption rate of calcium. [1, p.104]

Using the factorial method to convert the first and third rows of Tables 4.1 and 4.2 (input as the retention) to intake values, yields Table 4.3 (all in mg/d).

Table 4.3.: Factorial Method EAR and RDA

	Our Material		Bailey Material		
	males	females	males	females	IOM
EAR	1484	1273	1708	1466	1100
RDA	1835	1482	2140	1772	1300

As two examples, the following illustrate the EAR using our material for males and females respectively: the $1484 = \frac{273.7+290}{.38}$ and the $1273 = \frac{210.83+273}{.38}$.

RDA used the standard deviation via the Delta method and was compared to the value from the simulated standard deviation. These were nearly identical. The simulated version was used as a check and will be omitted moving forward.

Looking at the table, all 6 of the values from our work are above the current DRI values, with the males being even higher than the females. Furthermore, the predicted RDA values are even further from the recommendations than their EAR counterparts. The reason for this is because the CDC currently uses 100 mg/d as the standard deviation of calcium intake needed, so they add 200 mg/d to their EAR to get the RDA. However, while accounting for absorption percentage, about 351 and 209 mg/d are being added for the two genders. These are both above 200, hence the predicted RDA values are further away from the standards than the predicted EAR values. The values from Bailey et al. are above both the current DRI values and our values.

4.1.2 Papers from Weaver et al.

The remaining part of this chapter deals with different papers written by Connie Weaver et al. and the formulas relating calcium intake to calcium balance. Again, the formulas are being inverted to serve the need of the prediction of intake from balance. In these papers, balance is defined to be intake minus fecal and urinary excretion. Sweat losses are not accounted for. As defined previously, $g_p = b - s$ or $b = g_p + s$. In a 2005 paper by Weaver et al., they stated that calcium lost through sweat is about 54 and 51 mg/d in blacks and whites respectively. [9] Therefore, the balance values used in the following equations will be those in table 4.1 + 51 for the white estimates and + 54 for the black estimates.

First,"Racial differences in skeletal calcium retention in adolescent females with varied controlled calcium intakes" [10] was examined. The model of balance in blacks = 194.26 + .247 intake vs. balance = 9.75 + .247 intake in whites. Inverting these, gets

$$intake = \frac{balance - 194.26}{.247}$$

in black children and

$$intake = \frac{balance - 9.75}{.247}$$

in white children. Using our values for balance yields table 4.4 (all in mg/d). As an example, the EAR for white males is

$$intake = \frac{324.77 - 9.75}{.247} = 1275.$$

It is interesting that this model has an RDA for males well above the current value and the RDA for females is slightly above it. The males' EAR is above the current

Table 4.4.: Race Paper

		Males	Females
White	EAR	1275	1021
	RDA	1816	1342
Black	EAR	1288	1033
	RDA	1828	1354

value, but the females' EAR is below it.

Next, the model in the "Predictors of calcium Retention in adolescent males" paper was employed. [11] The model included was:

$$\begin{aligned} &balance \\ &= 657.6exp(-3.04 \\ &\quad *(1-.00112Intake)/(1+exp(-3.04*(1-.00112Intake))+144*(logIGF1-5.83). \end{aligned} \tag{4.2}$$

In the paper, values of 200, 350, and 600 were used for IGF levels (these had z-scores of -1, 0 , and 2 for the sample). Let

$$y = \frac{balance - 144(logIGF1 - 5.83)}{657.6}.$$

Then,

$$intake = \frac{1 + log(y/(1-y))/3.04}{.00112}$$

or

$$intake \approx 892.86 + 293.70log(y/(1-y)).$$

This model needs retention and IGF1. For now, the 3 IGF1 values in the paper (200, 350, and 600) are the ones used as inputs. As an example, the EAR estimate for white males (with IGF1 = 200) is shown in detail. For white males,

$$y = \frac{324.77 - 144 * (log(200) - 5.83)}{657.6} = .6103.$$

Plugging this into log(y/(1-y)) yields .4486. Finally,

$$intake = \frac{1 + .4486/3.04}{.00112} = 1025.$$

Table 4.5.: Adolescent Males Paper

	White Males		Black Males	
IGF1	EAR	RDA	EAR	RDA
200	1025	1325	1030	1334
350	878	1129	884	1135
600	736	979	742	984

Again, these are below those of the factorial method. This paper did not have an additional formula for females. Additionally, only the RDA when IGF1 = 200 was above the current value (1,300). None of the EAR values were over the current value (1,100).

The next paper examined was "Calcium retention in adolescent males on a range of controlled calcium intakes". [10] The model within is very similar to the previous.

$$balance = 525.3 * exp(-3.12(1-.00106Intake))(1+exp(-3.12(1-.00106Intake)))+171g,$$

where g is +.5 for males and -.5 for females. Let

$$y = \frac{balance \mp 85.5}{525.3}$$

(where the - is now for males). Then,

$$intake = \frac{1 + log(y/(1-y))/3.12}{.00106}$$

or,

$$intake \approx 943.40 + 302.37 log(y/(1-y)).$$

Again, the values for white males at EAR is used as an example. For them,

$$y = \frac{324.77 - 85.5}{525.3} = .4555,$$

and

$$log(y/(1-y)) = -.1785,$$

therefore

$$intake = \frac{1 - .1785/3.12}{.00106} = 889.$$

Table 4.6.: Calcium Requirements (in mg/d) Using the Range of Controlled Calcium Intakes Paper Formula

		Males	Females
White	EAR	889	1146
	RDA	1214	1387
Black	EAR	896	1153
	RDA	1222	1398

Just like previously, these values are below those from the factorial method. Interestingly, all male values are below the current standards while all female values are above the current standards.

The next paper was "Calcium retention in relation to calcium intake and postmenarcheal age in adolescent females". [12] The model was

$$balance = 565.5 * exp(-2.11 + .002intake)/(1 + exp(-2.11 + .002intake)) + 18.2 - 7.42PMA.$$

PMA is post menarcheal aqe in months. Note the average PMA for this study was about 10.7. Let $y = \frac{balance - 18.2 + 7.42PMA}{565.5}$. Then,

$$intake = \frac{log(y/(1-y)) + 2.11}{.002}$$

or

$$intake = 1055 + 500log(y/(1-y)).$$

This paper is only for females. It includes retention and PMA as inputs. PMA is calculated as 10.7 + (12.35 - 12.7)*12 = 6.5. The 12.35 is the Age of PBMCV for

females for our model. The 12.7 is the average age of females in the paper, and they said that the average value of PMA is 10.7 (in months). As an example, EAR for white females is examined. For them,

$$y = \frac{261.83 - 18.2 + 7.42 * 6.5}{565.5} = .5161,$$

then

$$log(y/(1-y)) = .0645,$$

and

$$intake = \frac{.0645 + 2.11}{.002} = 1087.$$

The remaining results are in table 4.7.

Table 4.7.: PMA Paper

		Females
White	EAR	1087
	RDA	1379
Black	EAR	1098
	RDA	1391

While the EAR values are slightly below the current values, the RDA values are both above the current amounts.

The last paper is "Obesity augments calcium-induced increases in skeletal calcium retention in adolescents." [13] They have two separate balance models based on the level of intake. When intake is > 1604.2, the model is

$$balance = 1030.4 + 1.95BMI - 113.9White + 105.4Male - 44.5Age.$$

Notice this model does not have intake in it, meaning the retention has leveled off. For intake < 1604.2,

$$\begin{aligned} balance = 782.6 - 1.78BMI + .1545Intake + .00232BMI \\ * Intake - 113.9White + 105.4Male - 44.5Age. \end{aligned} \tag{4.3}$$

For both of these formulas, age is measured to the nearest tenth of a year and BMI is the percentile of BMI. Both white and male are indicator variables. Inverting the formula yields

$$intake = \frac{balance - 782.6 + 1.78BMI + 113.9White - 105.4Male + 44.5Age}{.1545 + .00232BMI}.$$

For this formula to work, balance, age, and BMI are all needed. The ages of peak are 13.74 and 12.35 for males and females respectively. For percentiles of BMI, 0, 5, 85, and 95 were used. The rationale behind this is those values are the end points of the different groups. Underweight children fall in the 0-5 percentile range, normal weight children in the 5-85 percentile range, and the last two categories are overweight and obese respectively. Translating BMI values of 18, 25, and 35 to percentiles was also examined. (The results of these are the last 3 rows in the table.) The reason for these values is that they are typically associated with the borders between healthy, overweight, and (very) obese.

The percentiles come from the standard normal, Z. Z is calculated from the formula $Z = ((BMI/M)^L - 1)/(L * S)$. The values of the parameters (L, M, and S) are based on both age and gender. For males that are 13.71 years old, the parameter values are -2.26, 18.93, and .14 for L, M, and S respectively. These values are -2.25, 18.98, and .14 for males that are 13.79 years old. For females aged 12.29 vs. 12.38 years old the parameters are -1.97, 18.26, and .15 vs. -1.96, 18.31, and .15 respectively. These ages were picked because they were on each end of the ages of PBMCV. The appropriate z values and percentiles for each of these age, gender, and BMI (25 or 35) combinations were calculated, then the two ages for that gender and BMI level were averaged to get the following results.

Table 4.8.: BMI values to percentiles

Gender	BMI	Percentile
Male	18	34.23
Male	25	93.62
Male	35	99.3
Female	18	45.71
Female	25	94.12
Female	35	99.58

Here are the results in mg/d:

Table 4.9.: Obesity Paper Male's Estimates

	White		Black	
BMI	EAR	RDA	EAR	RDA
0	1049	1913	331	1195
5	1030	1833	362	1165
50	928	1422	518	1012
85	891	1271	576	955
95	883	1239	588	944
18 converted	953	1524	479	1050
25 converted	884	1244	586	945
35 converted	880	1227	592	939

Table 4.10.: Obesity Paper Female's Estimates

	White		Black	
BMI	EAR	RDA	EAR	RDA
0	924	1438	206	720
5	913	1391	245	723
50	857	1150	447	740
85	836	1062	521	747
95	832	1044	536	748
18 converted	860	1165	434	739
25 converted	832	1045	535	748
35 converted	830	1036	542	748

These estimates are all below those from the factorial method for EAR. However, some of the RDA values are above those from the factorial method. The general

pattern is that the higher the BMI percentile, the less calcium needed to retain the same amount. However, this is the opposite of the pattern for EAR for black children. This has to do with the opposite signs in front of intake and the interaction between intake and BMI.

Two other percentiles of BMI of note are 78.02 and 17.74. These are the values for which the white males' and white females' RDA respectively achieve the current recommendation of 1,300 mg/d.

4.1.3 Calcium Retention to Intakes Needed Using Average TBBMCV

As outlined in section 3.3.5, the important values relating to average TBBMCV by age group and gender are in Table 4.11.

Table 4.11.: Average and SD of the Differences in F(Θ) in Ca mg/d

	$\frac{F(14)-F(9)}{14-9}$	$SD\left(\frac{F(14)-F(9)}{14-9}\right)$	$\frac{F(18)-F(14)}{18-14}$	$SD\left(\frac{F(18)-F(14)}{18-14}\right)$
Males	213.81	57.39	156.72	75.32
Females	184.08	32.76	60.30	36.25

The average values are used directly in the factorial method to calculate the EAR for calcium intake. Just like with the PBMCV material, the average TBBMCV + 2*SD(average TBBMCV) is used as the input in the factorial method to calculate the RDA for calcium intake. The last table is a summary of the DRI values by age and gender according to these

All of these values are above the current EAR and RDA for calcium intake except for the females aged 14 to 18. Additionally, the male EAR and RDA values are higher than their female counterparts.

Lastly, these values are smaller than those established using PBMCV.

Table 4.12.: Average TBBMCV DRI Values Summary

Age Range	Gender	EAR	RDA
9 to 14	Males	1326	1628
14 to 18	Males	1176	1572
9 to 14	Females	1203	1375
14 to 18	Females	877	1068

Lastly is a summary of the values for both the peak rate of growth and the average rate of growth (by age).

			IOM	Our Peak	Our Average
9-14	Female	EAR	1100	1274	1203
9-14	Male	EAR	1100	1484	1326
14-18	Female	EAR	1100	1274	876
14-18	Male	EAR	1100	1484	1176
9-14	Female	RDA	1300	1482	1375
9-14	Male	RDA	1300	1834	1628
14-18	Female	RDA	1300	1482	1068
14-18	Male	RDA	1300	1834	1572

4.2 Conclusions

Using the established factorial method in conjunction with values of peak bone growth yielded EAR and RDA values above the current DRI values. The male intake values needed were even farther from the current DRI values than their female counterparts. Additionally, intake values needed were calculated with average bone accrual by gender and age group. The age groups were [9,14) and [14,18] since these are the groups in the current IOM manual for calcium intakes. For the males aged 9 to 14, males aged 14 to 18, and the females aged 9 to 14, all intake needed values calculated were above the current DRI values. For females aged 14 to 18, their intake needed values were below the current DRI values. For the formulas built by Weaver et al., some of the intake values were above and some were below the current DRI values. It should be noted that the data set used to analyze these growth curves were from an observational study and not from subjects that were forced to consume the recommended calcium amount per day from the IOM. Many people get less than the

required amount, yet they have peak rates of growth that necessitate more than the current requirement. For this reason, the estimates produced within this chapter may error on the conservative side.

5. SENSITIVITY ANALYSIS

This chapter focuses on other models or extensions of the current model

5.1 Extending the model beyond gender

This chapter focuses on other models or extensions of the current model. For example, would incorporating race, height, weight, BMI, or anything else yield significant results? Or would a different model work better?

Work was done to incorporate the height, weight, and BMI (since race is not in the Bailey data set). Different models for including these variables like putting them in the logistic regression, having them be linear, centered linear, and centered quadratic were run with the original variables and them divided by age. These models were run for the whole data set and the data < 22 years of age. A model with height and BMI together, as well as all three together (weight too) were run.

The best of these is the generalized logistic for age with an additional component for a linear weight model. This is because it had the lowest s^2 value and the parameters for the logistic still seemed reasonable with respect to their interpretations. Table 6.1 is the information from the best extended model: a linear weight in addition to the generalized logistic with age. The μ estimates are similar, the β and γ values have decreased, and the σ^{-1} values have increased compared to their counterparts from the asymmetric model (without weight). The positive coefficient for weight compensates for the smaller β.

Table 5.1.: Parameters for Generalized Logistic Regression on Age with a Linear Weight Component

	$\hat{\mu}$	$\hat{\beta}$	$\hat{\sigma^{-1}}$	$\hat{\gamma}$	$\hat{Weight}$
Male	16.101	0.85	11.46	0.03	0.03
Female	14.32	0.77	9.86	0.02	0.02

In addition to models with various explanatory variables added, different functions were also considered. Some were alternate versions of the logistic, most notably the symmetric version. Other functions used include the Weibull and Gompertz functions.

5.2 The Gompertz Distribution

5.2.1 Introduction to the Gompertz

The Gompertz Distribution has

$$F(x) = aexp(bexp(cx)).$$

The upper asymptote is a, b is the x-axis displacement, and c is a growth rate. Both b and c are negative numbers. It also has

$$f(x) = abcexp(cx)exp(bexp(cx)).$$

This can be rewritten as

$$f(x) = abcexp(bexp(cx) + cx).$$

Lastly,

$$f'(x) = abcexp(bexp(cx) + cx)(bcexp(cx) + c).$$

5.2.2 SAS results

Initial parameter estimates for the Gompertz model were provided by values from the asymmetric logistic model. Specifically, the values of β from the logistic models were used as a for the Gompertz model. The other 2 parameters were found solving the systems of equations (with age and value of PBMCV set equal to the logistic estimates) for each gender. Specifically, PBMCV was used first to solve for c (once a had been estimated). Then, the age of PBMCV was used to solve for b. This gave initial estimates for the SAS program. The programs were run on the same data set

as the logistic (namely, the Bailey data set with no ages over 22 years old). First, it was ran without the MVN assumption of the parameters. Then, the estimates from that output, including their standard deviations, were used as the initial guesses in the model with a MVN assumption. The result is below and was used to calculate the summary table.

The female parameter estimates were:

$$\hat{\Theta}_f = (\hat{a}_f, \hat{b}_f, \hat{c}_f) = (2.42, -11.96, -.26)$$

with a covariance structure of

$$\hat{\Sigma}_f = \begin{bmatrix} .145 & .292 & .004 \\ .292 & 8.533 & .108 \\ .004 & .108 & .002 \end{bmatrix}.$$

The male parameter estimates were:

$$\hat{\Theta}_m = (\hat{a}_m, \hat{b}_m, \hat{c}_m) = (5.08, -6.34, -.13)$$

with a covariance structure of

$$\hat{\Sigma}_m = \begin{bmatrix} 12.71 & 11.51 & .33 \\ 11.51 & 13.43 & .34 \\ .33 & .34 & .01 \end{bmatrix}.$$

5.2.3 Graphs and PBMCV Analysis

Figures 6.1 and 6.2 are graphs representing the fitted Gompertz models. They contain both f(x) (colored red) and F(x) (colored black. Figure 6.1 represents the male graphs, while Figure 6.2 represents the female graphs.

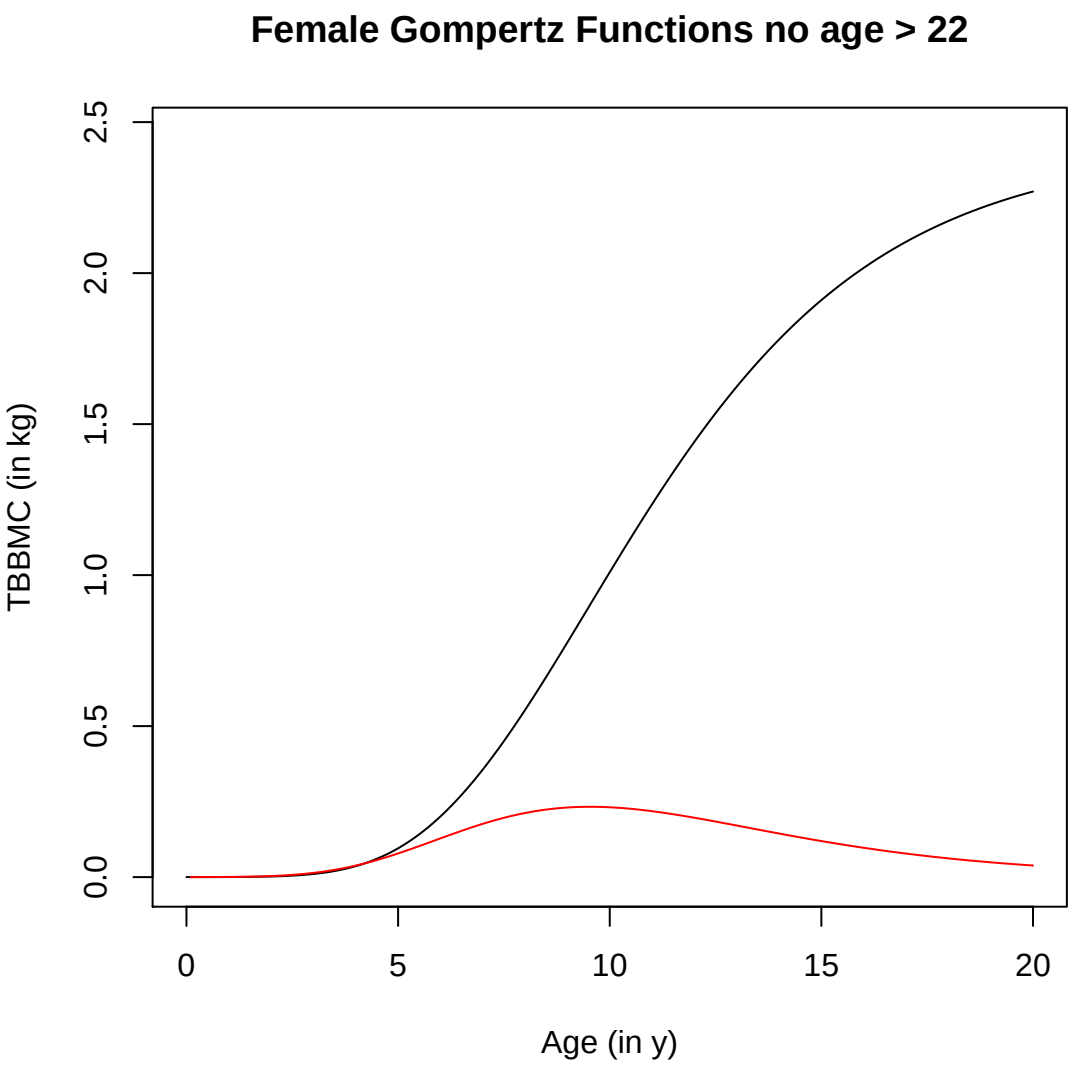
One difference between this model and the logistic is that the upper asymptote for the males is about 5.1 here compared to about 2.8 for the logistic model. However,

Figure 5.1.: Male Gompertz Distributions

the age range of interest is 9 to 18 years old. The values of the Gompertz and Logistic models are 2.76 and 2.73 kg respectively at 18 years old. The models are comparable for the ages desired. If the desired age range was 18 or older, the Gompertz model would not be realistic for the males. The male Gompertz function continues to grow until about age 46. Whereas, the logistic model as well as the female Gompertz model

level off around 20. Additionally, the peak for the Gompertz bone growth is flatter, meaning it sustains its peak growth (or close to it) for a longer period of time and has less variability in growth from year to year. For these reasons, the logistic model was preferred.

Figure 5.2.: Female Gompertz Distributions



5.2.4 Important Values from the Gompertz

The function for modeling TBBMC is:

$$F(x) = ae^{be^{cx}}.$$

Taking the derivative of F(x), yields f(x). Let u= be^{cx}. Then,

$$\frac{du}{dx} = bce^{cx}$$

and

$$f(x) = F(x)\frac{du}{dx} = abce^{ct}e^{be^{cx}} = abce^{be^{cx}+cx}.$$

The goal is to find the maximum of f(x), so its derivative is needed. This time, let u $= be^{cx} + cx$. Then,

$$\frac{du}{dx} = bce^{cx} + c$$

and

$$f'(x) = f(x)\frac{du}{dx} = abce^{be^{cx}+cx} * (bce^{cx} + c).$$

Set this equal to 0 and divide out the first part. What is left is $bce^{cx} + c = 0$. This becomes $x = \frac{ln(-1/b)}{c}$. This is the age of PBMCV. If this is plugged back into f(x), the c values cancel, and it is

$$abcexp(bexp(ln(-1/b)) + ln(-1/b))$$

which becomes

$$abcexp(-1 + ln(-1/b)) = abcexp(-1)(-b^{-1}) = -acexp(-1).$$

The variance of PBMCV is

$$\begin{aligned} Var(PBMCV) &= Var(-acexp(-1)) \\ &= exp(-2)Var(ac)(-2)(c^2Var(a) + a^2Var(c) + 2acCov(a,c)). \end{aligned} \tag{5.1}$$

Below is a table that represents these important values from the Gompertz model. The age of PBMCV is similar for males (at 13.67 compared to 13.74 from the logistic

model). However, this value for females is really low. In Bailey's papers, they had between 1 and 1.5 years difference across the genders. Here, the females value is just above 3 years different (and is about 3 yrs lower than what was estimated with the logistic, at 12.35). The opposite can be said for values of PBMCV. The females is close to its logistic counterpart (238.98 g/yr). The males value is about 70 below its logistic counterpart (310.33 g/yr). Both standard deviations are lower than their logistic counterparts. The females standard deviation is comparable (44.29 to 45.63). The males went down from 77.80 to 53.81 from the logistic to the Gompertz model. In summary, about half of the important values are similar between the 2 models (2/3 similar for females, and only 1/3 similar for males).

Table 5.2.: Age, Value, and Variance of PBMCV for the Gompertz Model

	Males	Females
Age PBMCV in yrs	13.67	9.49
PBMCV in g/yr	244.72	233.42
Var(PBMCV) in $(g/yr)^2$	2895.53	1961.69
SD(PBMCV) in g/yr	53.81	44.29
EAR in mg/d	1331	1260
RDA in mg/d	1581	1466

It turns out that the female values of EAR and RDA are not much different (off by 13 and 19 mg/d from the logistic model). The values for the males are off by much more (153 and 364 mg/d). It is interesting to see that the results for females are fairly robust. Lastly, all of these values are still above the current RDIs of 1,100 and 1,300 mg/d for EAR and RDA.

In conclusion, the other models examined were not an improvement upon the asymmetric logistic model. There were various reasons for this statement. Reasons

include rejecting normality (for the parameters, for the important values, for the residuals, or a combination thereof), having unrealistic parameter values, not being able to estimate a covariance structure and hence an associated RDA intake value, or a worse fit of the data.

6. SUMMARY AND FUTURE WORK

The asymmetric logistic model using age and gender provides a good estimate of bone accumulation. In addition, the values of peak and average bone growth are both Normal with estimable standard deviations. This means that they can be used to calculate both EAR and RDA values for calcium. Using the established concept of the factorial method to relate calcium retention and intake, the intake values need according to the peak rate of bone growth are all above the current recommendations. The male values are about 1,500 and 1,800 calcium mg/d for the EAR and RDA respectively based on the peak (compared to the current values of 1,100 and 1,300). The female values are about 1,300 and 1,500 calcium mg/d for the EAR and RDA respectively based on the peak (compared to the current values of 1,100 and 1,300).

Additionally, intake values needed were calculated with average bone accrual by gender and age group. The age groups were [9,14) and [14,18] since these are the groups in the current IOM manual for calcium intakes. For the males aged 9 to 14, males aged 14 to 18, and the females aged 9 to 14, all intake needed values calculated were above the current DRI values. For females aged 14 to 18, their intake needed values were below the current DRI values.

The next step is to extend the model beyond gender; in particular, an extension to include race. Another data set would need to be included to incorporate race. Additionally, further investigation of variables such as height, weight, and BMI will be conducted. Currently, these last 3 variables have been investigated; however, these analyses were run with the value of the variable and were found to not be an improvement over the current model. It might prove valuable to use percentiles of these variables (based at least on age). That way the correlation between these variables

and age can be removed or diminished. Once this is done, the next step is establishing formulas for relating calcium intake and calcium retention based on age as well as the additional variables.

REFERENCES

REFERENCES

[1] Institute of Medicine (US) Standing Committee on the Scientific Evaluation of Dietary Reference Intakes et al. *Dietary Reference Intakes*, 1997.

[2] Hassanali Vatanparast, Donald A Bailey, Adam DG Baxter-Jones, and Susan J Whiting. Calcium requirements for bone growth in canadian boys and girls during adolescence. *British journal of nutrition*, 103(04):575–580, 2010.

[3] Heather B Del Valle, Ann L Yaktine, Christine L Taylor, A Catharine Ross, et al. *Dietary reference intakes for calcium and vitamin D*, 2011.

[4] D. A. Bailey, H. A. Mckay, R. L. Mirwald, P. R. E. Crocker, and R. A. Faulkner. A six-year longitudinal study of the relationship of physical activity to bone mineral accrual in growing children: The university of saskatchewan bone mineral accrual study. *Journal of Bone and Mineral Research*, 14(10):1672–1679, 1999.

[5] Hassanali Vatanparast, Donald A Bailey, Adam DG Baxter-Jones, and Susan J Whiting. Calcium requirements for bone growth in canadian boys and girls during adolescence. *British journal of nutrition*, 103(04):575–580, 2010.

[6] George Casella and Roger L Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

[7] Donald A. Bailey, Alan D. Martin, Heather A. McKay, Susan Whiting, and Robert Mirwald. Calcium accretion in girls and boys during puberty: A longitudinal analysis. *Journal of Bone and Mineral Research*, 15(11):2245–2250, 2000.

[8] Anirban DasGupta. *Asymptotic theory of statistics and probability*. Springer Science & Business Media, 2008.

[9] Karin Wigertz, Cristina Palacios, Lisa A Jackman, Berdine R Martin, Linda Doyle McCabe, George P McCabe, Munro Peacock, J Howard Pratt, and Connie M Weaver. Racial differences in calcium retention in response to dietary salt in adolescent girls. *The American journal of clinical nutrition*, 81(4):845–850, 2005.

[10] Michelle Braun, Cristina Palacios, Karin Wigertz, Lisa A Jackman, Rebecca J Bryant, Linda D McCabe, Berdine R Martin, George P McCabe, Munro Peacock, and Connie M Weaver. Racial differences in skeletal calcium retention in adolescent girls with varied controlled calcium intakes. *The American journal of clinical nutrition*, 85(6):1657–1663, 2007.

[11] Kathleen M Hill, Michelle Braun, Mark Kern, Berdine R Martin, James W Navalta, Darlene A Sedlock, Linda McCabe, George P McCabe, Munro Peacock, and Connie M Weaver. Predictors of calcium retention in adolescent boys. *The Journal of Clinical Endocrinology & Metabolism*, 93(12):4743–4748, 2008.

[12] LA Jackman, Stephanie S Millane, Berdine R Martin, OB Wood, GP McCabe, M Peacock, and CM Weaver. Calcium retention in relation to calcium intake and postmenarcheal age in adolescent females. *The American journal of clinical nutrition*, 66(2):327–333, 1997.

[13] Kathleen M Hill, Michelle M Braun, Kara A Egan, Berdine R Martin, Linda D McCabe, Munro Peacock, George P McCabe, and Connie M Weaver. Obesity augments calcium-induced increases in skeletal calcium retention in adolescents. *The Journal of Clinical Endocrinology & Metabolism*, 96(7):2171–2177, 2011.

[14] Dorothy Teegarden, William R Proulx, Berdine R Martin, Jian Zhao, George P Mccabe, Roseann M Lyle, Munro Peacock, Charles Slemenda, Conrad C Johnston, and Connie M Weaver. Peak bone mass in young women. *Journal of Bone and Mineral Research*, 10(5):711–715, 1995.

[15] AC Looker, LG Borrud, JP Hughes, B Fan, JA Shepherd, and M Sherman. Total body bone area, bone mineral content, and bone mineral density for individuals aged 8 years and over: United states, 1999-2006. *Vital and health statistics. Series 11, Data from the national health survey*, (253):1–86, 2013.

Appendix A: Factorial Method Explained

The factorial method was referenced, which is a formula for relating the calcium retained to intake. It contains information about retention, intake, losses, and percent absorption. For this part, assume that TBBMCV is in grams per year. TBBMCV is how much the bones grew. This can be changed into how much calcium was retained. The factorial method formula is

$$Ca \quad intake = \frac{TBBMCV * \frac{1000}{365} * .322 + k}{.38}$$

The $\frac{1000}{365}$ converts from g/yr to mg/d. The 32.2% is the % of bone attributed to calcium. [7] The letter k is used for total calcium losses. It is divided into urinary (127 and 106 mg/d) [14], endogenous fecal (108 and 112 mg/d), and sweat losses (55 mg/d each). So, the k values are 290 and 273 mg/d for males and females respectively [1, p.104] The 38% is for calcium absorption. [1, p.104]

Below are some tables involving the factorial method to relate the calcium retention values from our model to calcium intake values.

Table A.1.: Average TBBMCV Per Age Range as an Input to the Factorial Method, EAR Version

Age/ Gender	Average Calcium Accretion (mg/day)	Urinary Losses (mg/day)	Endogenous Fecal Calcium Losses (mg/day)	Sweat Losses (mg/day)	Total Needed (mg/day)	Absorption %	Estimated Total Intake (Adjusted for Absorption) (mg/day)
9-13 Female	184	106	112	55	457	38	1203
9-13 Male	214	127	108	55	504	38	1326
14-18 Female	60	106	112	55	333	38	877
14-18 Male	157	127	108	55	447	38	1176

Table A.2.: Average TBBMCV Per Age Range as an Input to the Factorial Method, RDA Version

Age/ Gender	Average Calcium Accretion (mg/day)	Urinary Losses (mg/day)	Endogenous Fecal Calcium Losses (mg/day)	Sweat Losses (mg/day)	Total Needed (mg/day)	Absorption %	Estimated Total Intake (Adjusted for Absorption) (mg/day)
9-13 Female	228	106	112	55	501	38	1319
9-13 Male	303	127	108	55	593	38	1561
14-18 Female	152	106	112	55	425	38	1117
14-18 Male	314	127	108	55	604	38	1591

Table A.3.: PBMCV as an Input to the Factorial Method

Requirement/ Gender	Average Calcium Accretion (mg/day)	Urinary Losses (mg/day)	Endogenous Fecal Calcium Losses (mg/day)	Sweat Losses (mg/day)	Total Needed (mg/day)	Absorption %	Estimated Total Intake (Adjusted for Absorption) (mg/day)
EAR Female	211	106	112	55	484	38	1273
EAR Male	274	127	108	55	564	38	1484
RDA Female	290	106	112	55	563	38	1482
RDA Male	407	127	108	55	697	38	1835

Appendix B: Residual Analysis

B.5 r^2 and ANOVA material

This part examines residuals for the 3 different models with $\Theta = (\mu, \sigma^{-1}, \beta, \gamma = .2)$. The 3 models pertaining to individuals, full model, and full model with gender stratification. As a quick note, the totals column is the sum of the other two columns by construction. However, none of the columns have the property where SSR + SSE = SST. This is due to the fact that $\sum_i (Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})$ is $\neq 0$ (while it is $= 0$ within the linear regression framework). For examples, in the individual parameterization model, the men's SST = SSR + SSE - 1.63, and the women's SST = SSR + SSE - .83. Consequently, for the whole data set, SST = SSR + SSE - 2.46. While these values are close to being the sum, in the full model they are much further apart, and the full model with gender stratification is in between the two (with a moderate difference).

Below are 5 tables with Sum of Squares (SS) for each gender according to the regression model ($\sum_i (\hat{Y}_i - \bar{Y})^2$), error ($\sum_i (Y_i - \hat{Y}_i)^2$), and total ($\sum_i (Y_i - \bar{Y})^2$). These relate to the columns. The column labeled "Total" is just the sum of the gender columns. This is used to assess an overall fit of the model. Some information that is useful to the degrees of freedom:

- the individual data set has 1,350 observations on 188 people.
- the number of parameters is 188*3 + 1 = 565. The 3 is for (μ, β, σ^{-1}) for each person and the one refers to γ which is set at .2 for everyone.
- the number of parameters in the other 2 models are 4 and 7 respectively.
- The number of observations is 1,518 on 251 people for the whole data set.

Additionally, after each of the 5 tables of SS, there will be an ANOVA table complete with a pseudo-r^2 value. This value is 1 - $\frac{SSE}{SST}$. The reason for calling it a pseudo value is because this is one of the ways to calculate r^2 in a regression or ANOVA framework, but the relation that SSR + SSE = SST does not exist here.

This part does not use the other potential formula, namely, $r^2 = \frac{SSR}{SST}$ since this value was > 1 in 2 of the 3 columns for the individual parameterization model, which by definition is impossible. (It is almost > 1 for the females too.)

Table B.1.: SS table for the Individual Parameterization Model

	Males	Females	Total
Regression	197.54	119.65	317.19
Error	1.42	0.84	2.26
Total	197.33	119.67	317.00

Table B.2.: ANOVA table for the Individual Parameterization Model

	SS	df	MS	F	p-value	pseudo-r^2
Regression	317.19	565	0.56	194.81	0	0.992873
Error	2.26	784	0.00			
Total	317.00	1349				

Table B.3.: SS table for the 1 overall Θ Parameterization Model

	Males	Females	Total
Regression	141.56	204.29	345.85
Error	114.34	84.72	199.05
Total	357.84	228.54	586.39

Tables B.3 - B.6 were summaries of these models on the individual parameterization data set. They are used as a comparison of tables B.1 and B.2. For the 3 models, the model helps explain about 99.29%, 66.05%, and 73.77% of the variation in TBBMC. Not surprisingly, the individual parameterization does the best and the

Table B.4.: ANOVA table for the 1 overall Θ Parameterization Model

	SS	df	MS	F	p-value	pseudo-r^2
Regression	345.85	4	86.46	584.23	3.2E-292	0.660542
Error	199.05	1345	0.15			
Total	586.39	1349				

Table B.5.: SS table for the Θ by Gender Parameterization Model

	Males	Females	Total
Regression	236.10	146.16	382.26
Error	72.76	70.44	143.19
Total	335.72	210.15	545.87

Table B.6.: ANOVA table for the Θ by Gender Parameterization Model

	SS	df	MS	F	p-value	pseudo-r^2
Regression	382.26	7	54.61	511.80	0	0.737681
Error	143.19	1342	0.11			
Total	545.87	1349				

gender stratified model is better than the unstratified model. There is more variability in the males models than in the females models even though there are only 613 male observations compared to 737 female observations for this data set. In all 3 of these models, the p-value for the F statistic is about 0, so one would conclude that the model provides a better fit than just the mean of TBBMC.

Another test that could be of interest, is a type of nested models comparison. With this one could test whether having a different Θ according to gender is statistically

significant. In this context, H_0: 1 Θ is adequate (vs. a Θ for each gender) vs. H_A: 1 Θ is not adequate (vs. a Θ for each gender). The test statistic is,

$$F = \frac{\Delta SSE}{\Delta p} / MSE_{full}.$$

Here, this is

$$F = \frac{199.05 - 143.19}{7 - 4} / .11 = 174.51.$$

The degrees of freedom are 3 and 1342 and the p-value is about $1.60{*}10^{-95}$. It can be concluded that the stratified model provides a better fit than the unstratified model. The tables for the whole data set, which will be tables B.7 - B.10, are what is of primary concern.

Table B.7.: SS table for the 1 overall Θ Parameterization Model for the whole data set

	Males	Females	Total
Regression	157.12	218.63	375.74
Error	122.49	92.86	215.35
Total	384.73	246.11	630.84

Table B.8.: ANOVA table for the 1 overall Θ Parameterization Model for the whole data set

	SS	df	MS	F	p-value	pseudo-r^2
Regression	375.74	4	93.94	659.97	0	0.658624
Error	215.35	1513	0.14			
Total	630.84	1517				

There is more variability in these than with the first data set (in both SS_{error} and SS_{total}). However, this is to be expected with more observations and more people

Table B.9.: SS table for the Θ by Gender Parameterization Model for the whole data set

	Males	Females	Total
Regression	261.61	157.90	419.51
Error	83.14	78.94	162.07
Total	368.03	232.12	600.15

Table B.10.: ANOVA table for the Θ by Gender Parameterization Model for the whole data set

	SS	df	MS	F	p-value	pseudo-r^2
Regression	419.51	7	59.93	558.35	0	0.729943
Error	162.07	1510	0.11			
Total	600.15	1517				

included (168 and 63 more respectively). Both of these models are a significant improvement over the mean in predicting TBBMC. Lastly, the nested models statistic is

$$F = \frac{215.35 - 162.07}{3} / .11 = 165.46.$$

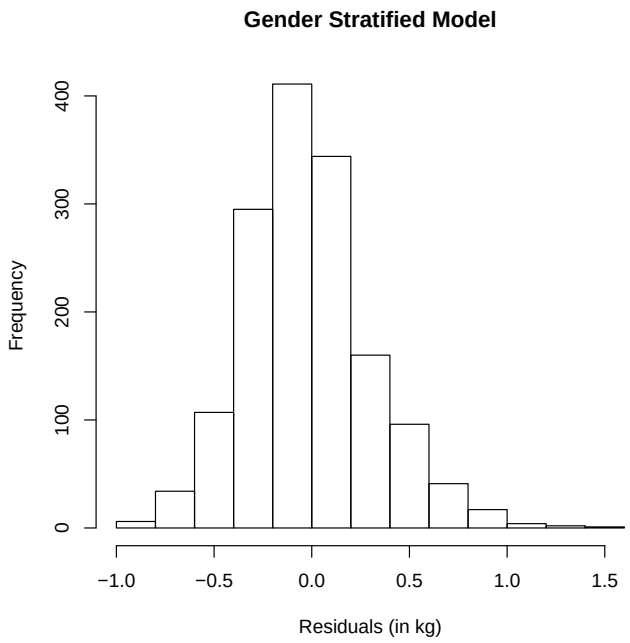
This has degrees of freedom 3 and 1,510 for a p-value of $9.89{*}10^{-93}$. The stratified model is significantly better than the unstratified model.

B.6 Histograms, QQplots, and Normality tests

Unlike the previous section, this part will only have information pertaining to the gender stratified model on the whole data set, which is the final model.

The overall histogram looks close to symmetric with perhaps a slight right-skewness.

Figure B.1.: Residual Histogram for Final Model



In the age versus residual plot, the mean appears to be about 0. There might be evidence of a slight heteroskedasticity.

The male histogram looks great. It is centered around 0, extends in both directions about 1 unit, and seems to be symmetric.

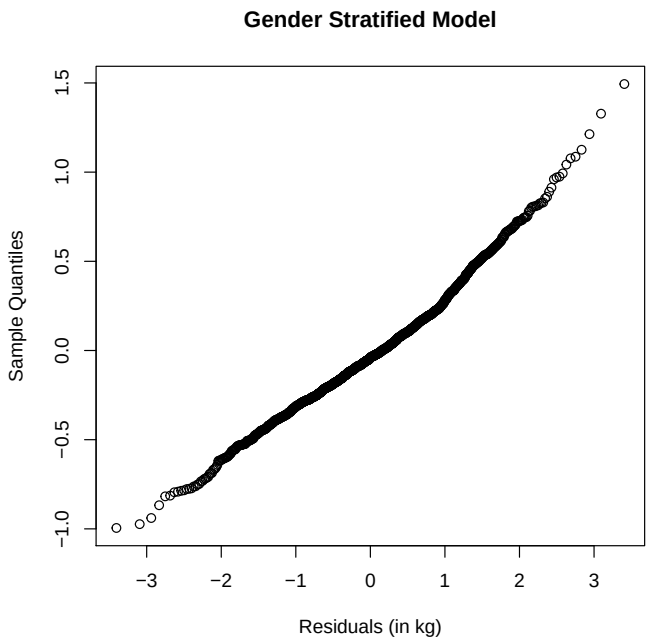
The male QQplot also appears to be good, with it being almost a 45 degree line.

The female histogram seems to have a mean around 0, but it looks right-skewed. It only goes down about .6 units, but up about 1.5 units.

The female QQplot looks like the line is sagging under a 45 degree angle. Again, this reinforces the right-skewness seen in the histogram.

This is an age versus residual plot where the female subjects are identified by the red color (males are the black circles). In figures B.9 and B.10 there are 2 vertical lines, one red and one blue. The red line refers to the age of PBMCV provided by the model and the blue line is the sample mean of age for that gender. In figure B.9,

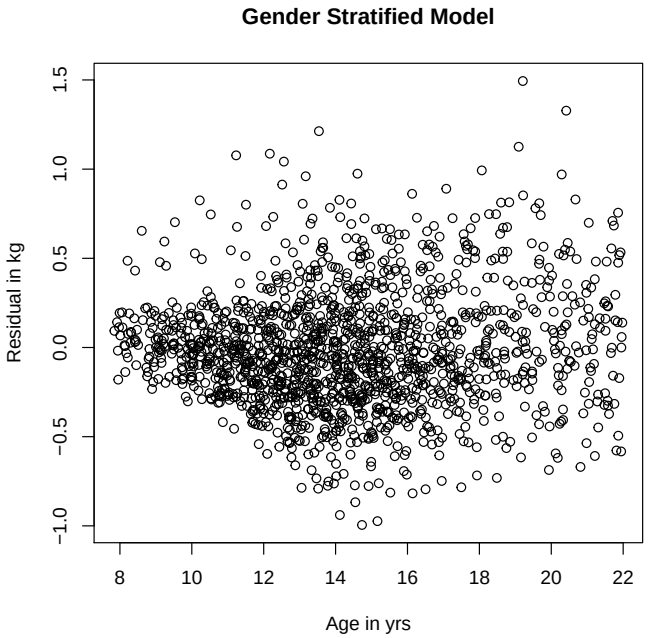
Figure B.2.: Residual QQplot for Final Model



the red line is at 13.74 and the blue line is at 14.47. In figure B.10, these lines are at 12.35 and 14.36.

The heteroskedasticity seen before is the result of the male residuals, not the females.

Figure B.3.: Age vs. Residual for Final Model



Appendix C: BMI

This appendix contains notes about converting BMI to percentiles.

$$Z = ((BMI/M)^L - 1)/(L * S)$$

where M, L, and S are age and gender dependent, and Z refers to a Standard Normal random variable. The idea is to convert to Z, then use the CDF to find the percentile. The necessary parameter values were found on:

`http://www.cdc.gov/growthcharts/percentile_data_files.htm`

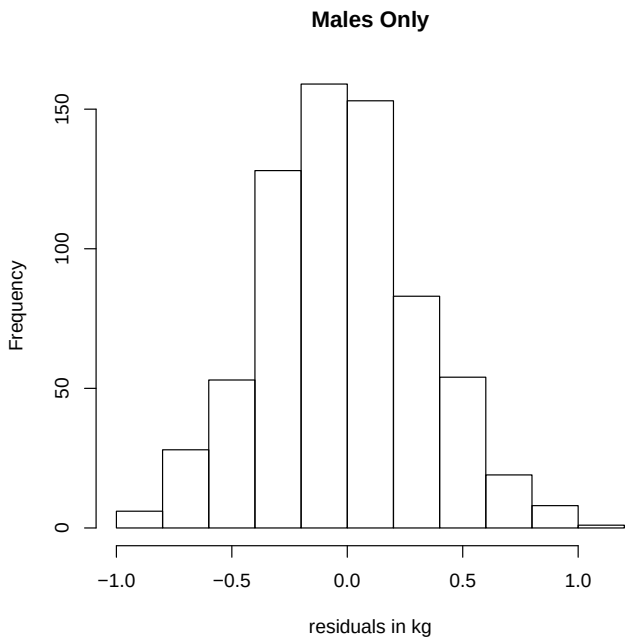
which is a CDC web site. However, that previous formula is for when L $\neq$ 0. If L = 0, then they use BMI = M exp(SZ) or Z = ln(BMI/M)/S.

The CDC gives the third, fifth, tenth, twenty-fifth, fiftieth, seventy-fifth, eighty-fifth, ninetieth, ninety-fifth, and ninety-seventh percentiles for BMI (along with gender, age, and parameter values).

Table C.1.: BMI values to percentiles

Gender	BMI	Percentile
Male	25	93.62
Male	35	99.3
Female	25	94.12
Female	35	99.58

Figure B.4.: Male Histogram for Final Model

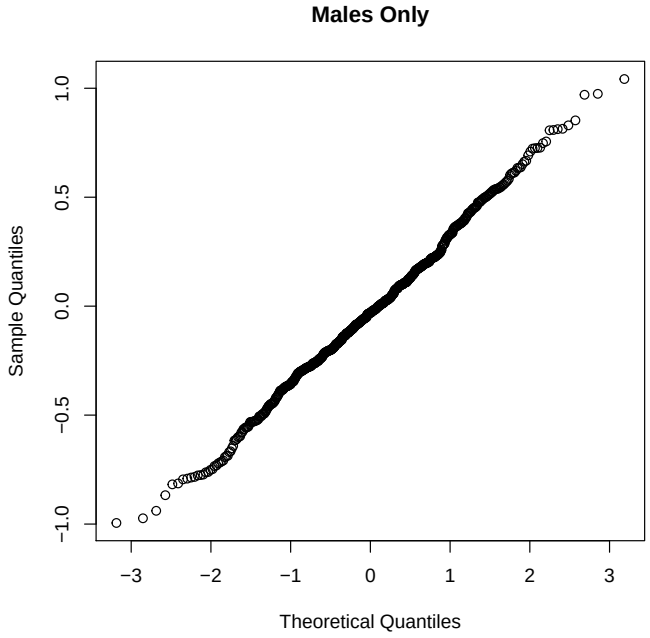


The ages of PBMCV are 13.74 and 12.35 for males and females respectively. The table above from the CDC has age in half month increments, so the two values on each side of the aforementioned ages (13.74 and 12.35) were selected. The previous table is the average for those two different time values by gender. The following table shows a little more detail of the calculations including the L, M, and S parameter values.

Table C.2.: BMI values to percentiles, detailed

Gender	BMI value	Age (in y)	L	M	S	Z	Percentile
males	25	13.71	-2.26	18.93	0.14	1.53	93.69%
males	25	13.79	-2.25	18.98	0.14	1.52	93.55%
males	35	13.71	-2.26	18.93	0.14	2.46	99.31%
males	35	13.79	-2.25	18.98	0.14	2.46	99.30%
females	25	12.29	-1.97	18.26	0.15	1.57	94.20%
females	25	12.38	-1.96	18.31	0.15	1.56	94.05%
females	35	12.29	-1.97	18.26	0.15	2.46	99.31%
females	35	12.38	-1.96	18.31	0.15	2.45	99.29%

Figure B.5.: Male QQplot for Final Model



Appendix D: Normal MGF Material

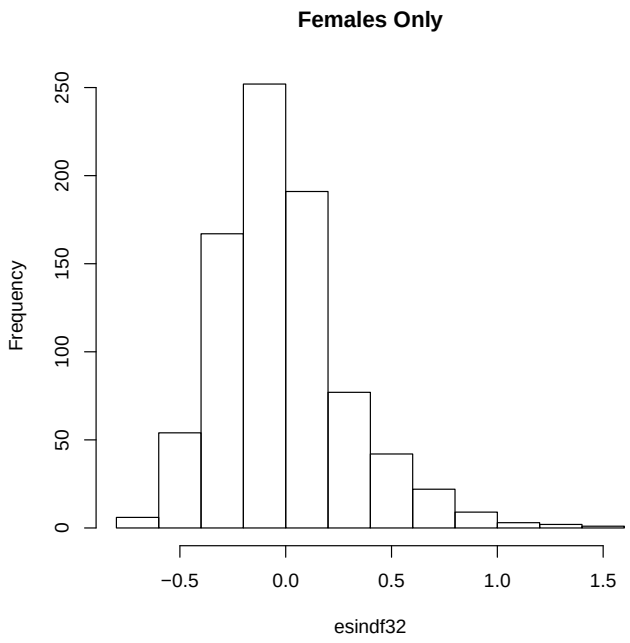
The following looks at expectations of a Normal Distribution to find some next level covariances.

To start, the moment generating function (mgf) of a Normal is

$$m(t) = exp(\mu t + .5\sigma^2 t^2).$$

$m'(t) = (\mu + \sigma^2 t)\mathrm{m(t)}$. So, $m'(0) = (\mu + 0)m(0) = \mu$ since $\mathrm{m(0)} = 1$.
$m''(t) = \sigma^2 m(t) + (\mu + \sigma^2 t)m'(t)$. So, $m''(0) = \sigma^2 * 1 + (\mu + 0) * \mu = \sigma^2 + \mu^2$.
$m'''(t) = \sigma^2 m'(t) + \mu m''(t) + \sigma^2 m'(t) + \sigma^2 t m''(t)$ which can be rewritten as:
$2\sigma^2 m'(t) + \mu m''(t) + \sigma^2 t m''(t)$. So, $m'''(0) = 2\sigma^2\mu + \mu(\sigma^2 + \mu^2) + 0 = \mu^3 + 3\mu\sigma^2$.
The reason to go this far is to be able to look at Cov(β, β^2) (or similar for σ^{-1}).

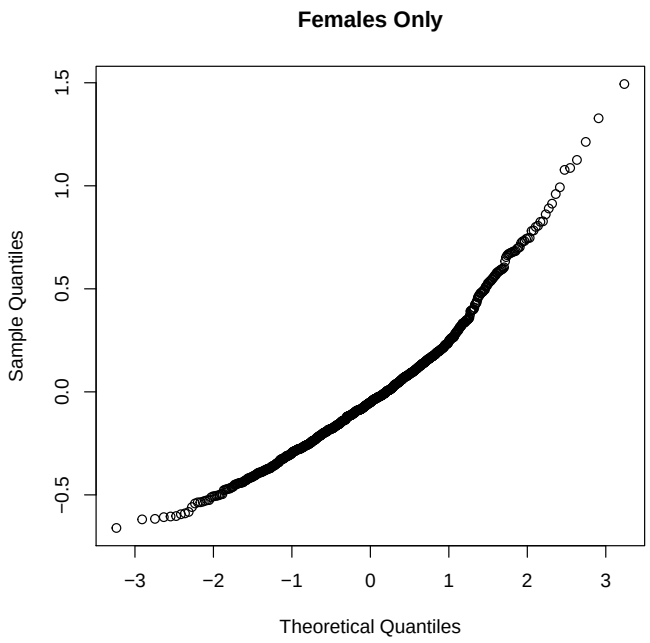
Figure B.6.: Female Histogram for Final Model



Cov(β, β^2) = E[β β^2] - E[β]E[β^2] = E[β^3] - E[β]E[β^2] = $\mu_\beta^3 + 3\mu_\beta\sigma_\beta^2$ - ($\mu_\beta\sigma_\beta^2 + \mu_\beta^3$) = $2\sigma_\beta^2\mu_\beta$.
Similar would apply to σ^{-1}.

Lastly, the goal is to find $m''''(t)$ to find $Var(X^2)$ (for example one could replace X with β or σ^{-1}).
$m''''(t) = 2\sigma^2 m''(t) + \mu m'''(t) + \sigma^2 m''(t) + \sigma^2 t m'''(t)$ which can also be rewritten. It is: $3\sigma^2 m''(t) + \mu m'''(t) + \sigma^2 t m'''(t)$.
Then, $m''''(0) = 3\sigma^2 * (\sigma^2 + \mu^2) + \mu(\mu^3 + 3\sigma^2\mu) + 0 = 3\sigma^4 + \mu^4 + 6\sigma^2\mu^2$.
Then, Var$(X^2) = E[X^4] - E[X^2]^2 = 3\sigma^4 + \mu^4 + 6\sigma^2\mu^2 - (\sigma^2 + \mu^2)^2$
$= 3\sigma^4 + \mu^4 + 6\sigma^2\mu^2 - (\sigma^4 + \mu^4 + 2\sigma^2\mu^2) = 2\sigma^4 + 4\sigma^2\mu^2$.

Figure B.7.: Female QQplot for Final Model



Appendix E: Data Exclusion Rules

E.7 Individual Parameter Estimates

When working with the individual subjects' parameterizations, the data is limited to subjects that had enough data to provide realistic parameter values. Only subjects with 5 or more data points were used. The purpose of modelling the individuals is to establish starting values within the other models as well as test assumptions (like normality). The following rules were initial guesses at good parameterizations:

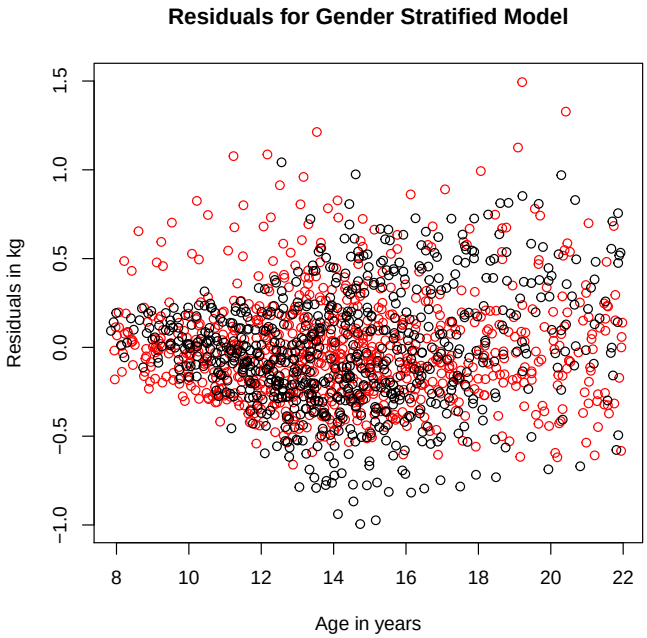
$$2 < \hat{\beta} < 4.5,$$

$$13 < \hat{\mu} < 20,$$

and

$$0 < \hat{\sigma^{-1}} < 2.$$

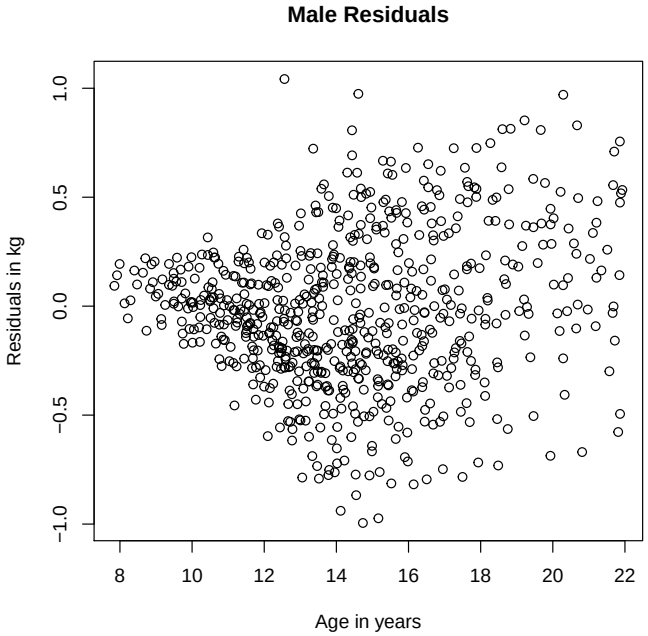
Figure B.8.: Age vs. Residual for Final Model with Red for Females



The first (β) rule makes the most biological sense alone. You can think of β as an estimate for "adult TBBMC", or how much TBBMC a person has after they have become an adult (stopped growing). The data set had a maximum TBBMC of about 3.8, so initially it allowed for this individual (or others) to have more growth after this point. The rule related to μ is a holdover from the symmetric model, when μ was the age of PBMCV. Now, $log(\gamma)/\sigma^{-1} + \mu$ is the age of PBMCV. The majority of this value comes from μ, so it can still be interpreted similarly. The above μ rule relates to about a range of 11.5 to 17.5 for the age of PBMCV.

The above rules were changed based on scatterplots and assessments of outlying values within those scatterplots, as well as the physical interpretations of the parameters. The value for $\hat{\beta}$ was changed to [2,4] for males and changed to [1.5,3.5] for females. The CDC has a manual studying individuals of different races and ages with measurements on TBBMC, TBBMD, etc. In it, they state that for 20-29 year old

Figure B.9.: Age vs. Residual for Final Model for Males only



males, the 5^{th} and 95^{th} percentiles of TBBMC are about 2.102 kg and 3.579 kg (1.704 and 2.692 for the 20-29 year old females). [15, p. 10] The rules used include the middle 90% from the CDC study. Since μ no longer had quite the same interpretation, it had a bit more flexibility. The new rules were $12 < \hat{\mu} < 19$ for males and $11 < \hat{\mu} < 18$ for females. The difference in genders here makes sense since females typically had their peaks about 1-2 years before males. Lastly, the rule for σ^{-1} was similar for both genders, namely $.6 < \hat{\sigma^{-1}} < 1.4$ for males and $.5 < \hat{\sigma^{-1}} < 1.4$ for females. Extreme values of $\hat{\sigma^{-1}}$ led to extreme values of PBMCV. Here is a summary table for the acceptable ranges of the parameters based on gender:

Examples of Individuals within the Data Rules Ranges

Below are 3 examples of individual estimates that are within the data rules ranges. Their fits are deemed to be good not only because the red line (the estimated logis-

Figure B.10.: Age vs. Residual for Final Model for Females only

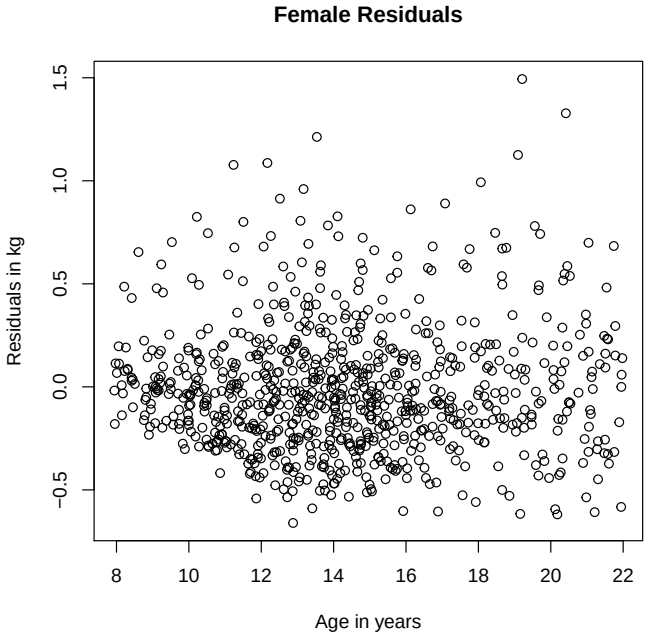


Table E.1.: Data Exclusion Rules

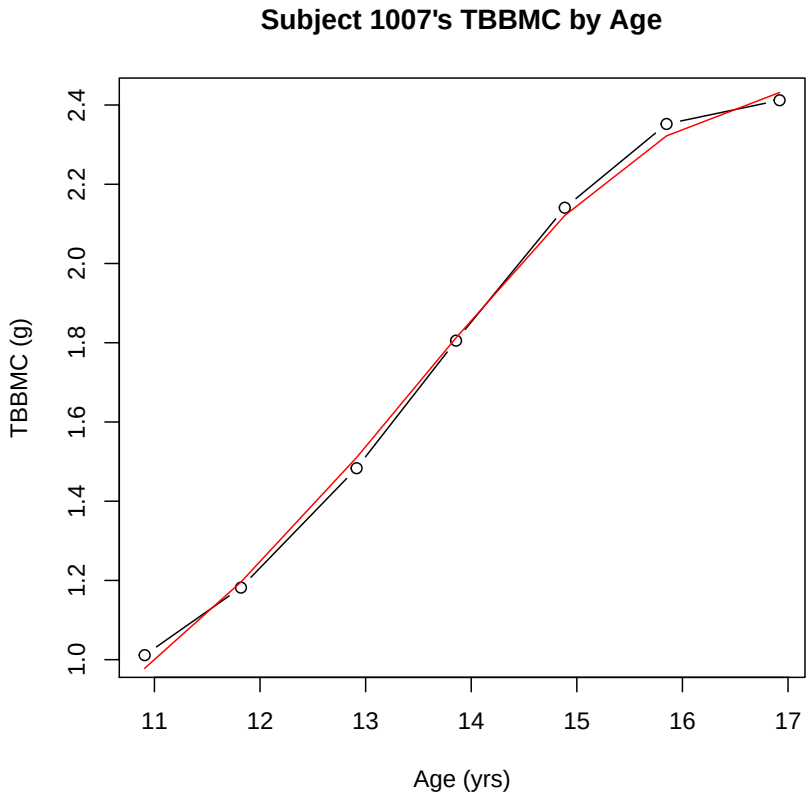
parameter	Males' range	Females' range
μ	[12,19]	[11,18]
β	[2,4]	[1.5,3.5]
σ^{-1}	[.6,1.4]	[.5,1.4]

tic curve) fit the data (the black line) well, but also because the parameters made conceptual sense since they have biological significance. Accompanying the graphs are the not only the Θ for these individuals but the ages and values of PBMCV to interpret the parameters.

ID 1007 had Θ = (15.07, 2.49, 1.12, .20) with age and values of PBMCV as 13.64 (y) and .33 (kg/y). One thing to notice about 1007's data is that their age of

PBMCV is around 14 and they have several values on each side of it. For this reason, the model not only provides a good fit, but also it has good biological interpretations. Additionally, the leveling off of TBBMC from age 16 to 17 allows for an easier fit of β.

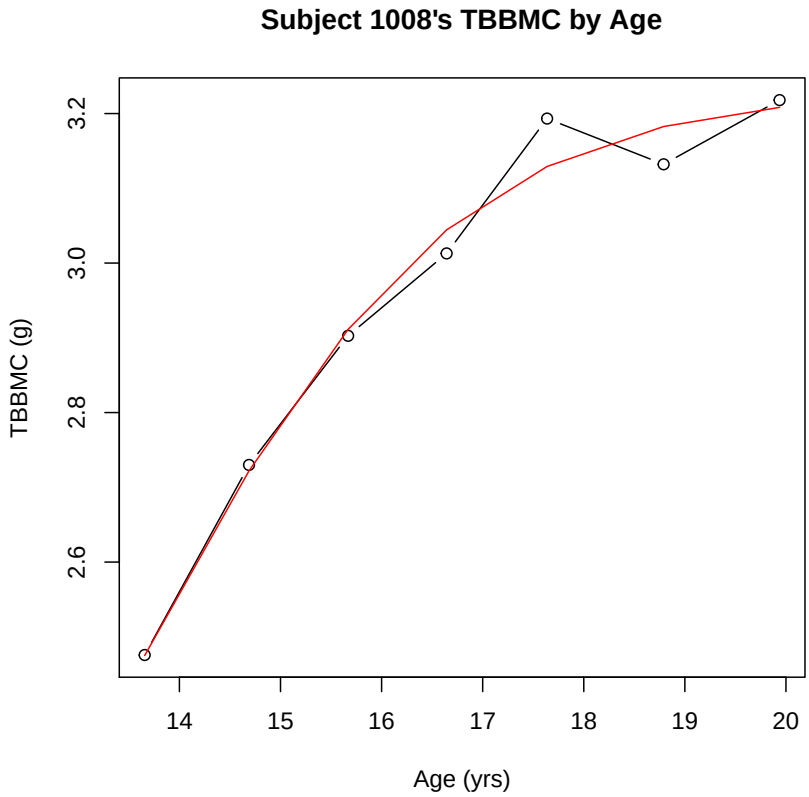
Figure E.1.: Person with Age of PBMCV Example 1



ID 1008 had $\Theta = (15.12, 3.23, .70, .20)$ with age and values of PBMCV as 12.80 (y) and .246 (kg/y). While 1008 had his age of PBMCV at around 13, all of his data is after that point. There was a noticeable leveling off of TBBMC from about 18 - 20 years of age. This allowed for a good estimation.

ID 2009 had $\Theta = (12.58, 2.24, 1.22, .20)$ with age and values of PBMCV as 11.26 (y) and .32 (kg/y). Just like 1007, ID 2009 had several data points on each side of

Figure E.2.: Person with Age of PBMCV Example 2

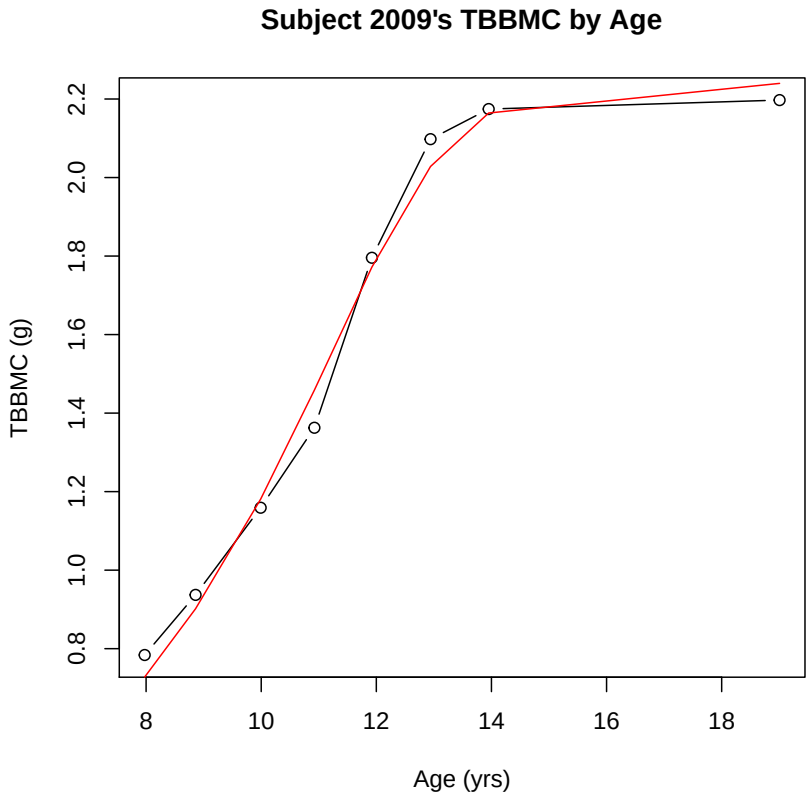


her age of PBMCV (11.26). Additionally, there is almost no change in TBBMC from 14 to 18 years old. These allowed for a good fit.

Examples of Individuals outside the Data Rules Ranges

Next are examples of 3 individual parameterizations that did not fall within the data rules ranges. These individuals had predicted curves that are not necessarily bad in the sense that the fit (red) does not accurately depict the data (black). However, their parameters do not make biological sense.

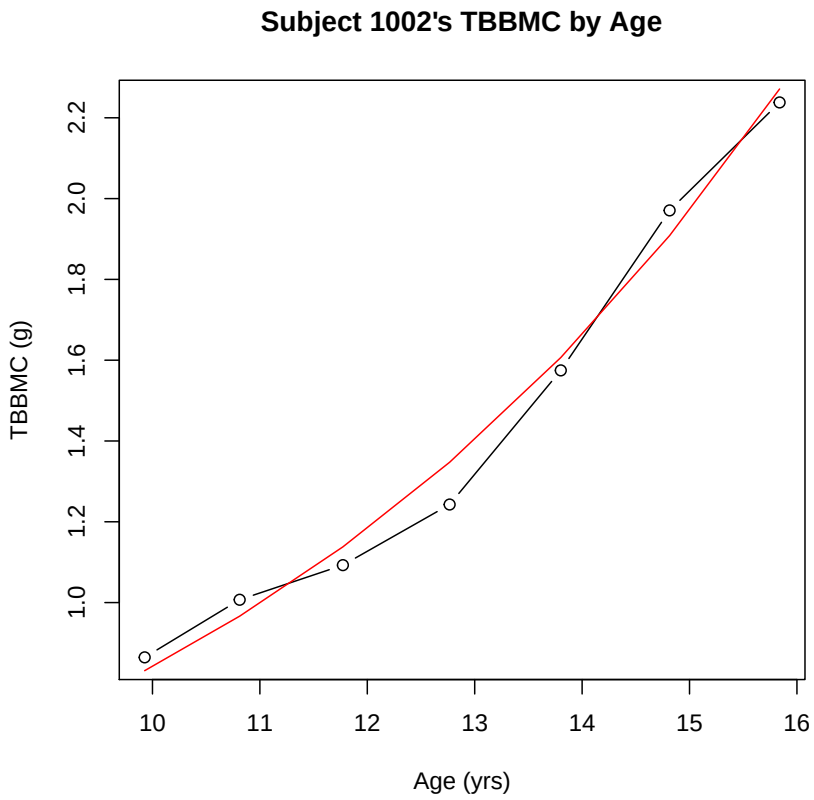
Figure E.3.: Person with Age of PBMCV Example 3



ID 1002 had $\Theta = (28.33, 18.99, .85, .20)$ with age and values of PBMCV as 26.44 (y) and 1.89 (kg/y). This male, along with the next 2 subjects had a very large $\hat{\mu}$, which in turn caused a large estimate of the age of PBMCV. His estimated age of PBMCV is around 26.5, but all of the data is before this point. One reason this estimate is unrealistic is that males typically have their age of PBMCV around 14. Furthermore, the value of PBMCV is 1.89, when this is reasonable between about .2 and .6. Why did the model have trouble fitting a reasonable curve? The data exhibits similar growth in TBBMC, and it appears to keep getting bigger throughout the time examined. This individual parameterization would benefit from another few observations after the data collected (just like the next 2 subjects). Then, there would be a decline in TBBMCV and it would start to level off, thus making it easier to fit a

curve, especially one with a more realistic age of PBMCV (which is most likely about 15-16 for this individual).

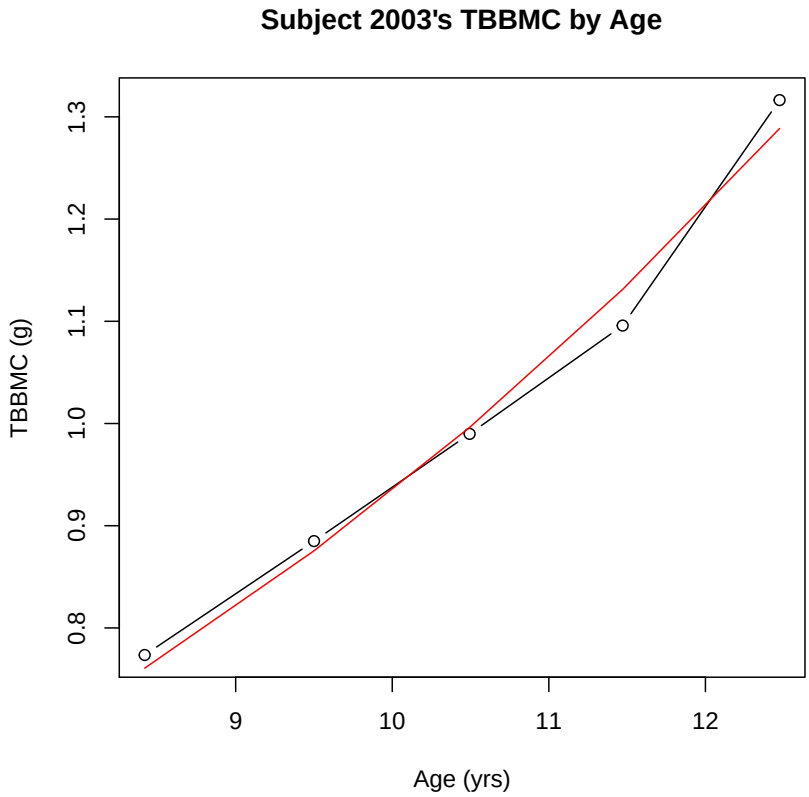
Figure E.4.: Person without Age of PBMCV Example 1



ID 2003 had $\Theta = (31.38, 15.05, .65, .20)$ with age and values of PBMCV as 28.90 (y) and 1.14 (kg/y). Again, ID 2003 exhibits a large value of age of PBMCV and all of her data is before this point.

ID 2004 had $\Theta = (29.31, 14.26, .70, .20)$ with age and values of PBMCV as 27.01 (y) and 1.16 (kg/y). Once again, ID 2004 exhibits a large value of age of PBMCV and all of her data is before this point.

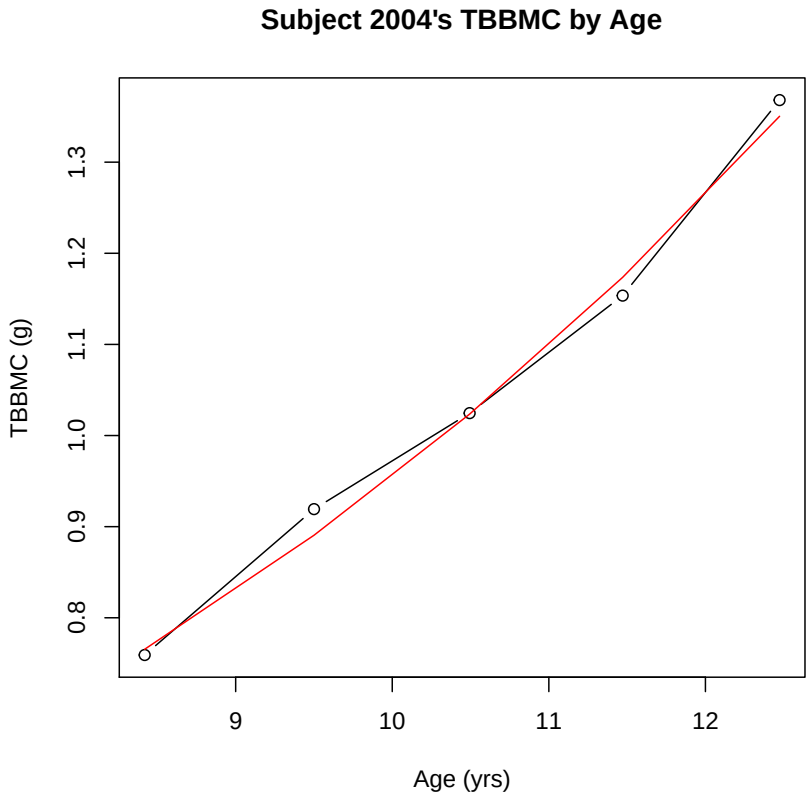
Figure E.5.: Person without Age of PBMCV Example 2



One common trait among the subjects with unrealistic parameters is that their data exhibit almost a straight line growth characteristic. Putting it another way, they were observed during a large (almost uniform) growth phase and there was little, if any, incline/decline in growth at the ends of the ages provided. There was no leveling off of the growth, so it was hard to fit an asymptote and an age of peak growth. This accounts for the unusual values of β and μ. In an ideal world, there would be more data at each end of their current data and would be better able to fit a model.

One key observation that should be stated is that there is nothing wrong with this data. If it is combined with other values on the other side of peak growth, it can be reasonably estimated. There is no reason to exclude it when looking at an overall

Figure E.6.: Person without Age of PBMCV Example 3



model.

E.8 Overall Model

While the aforementioned individuals with parameterizations outside the data rules ranges were included in the model, there were some constraints posed on the overall data set. This mainly stemmed from how the data was collected. Bailey et al. originally had a grant for a 6 year study on bone growth in adolescents, so most of the children have at least these 6 years worth of data. Later on, they received another grant and were able to follow-up with some of the individuals for a few more years.

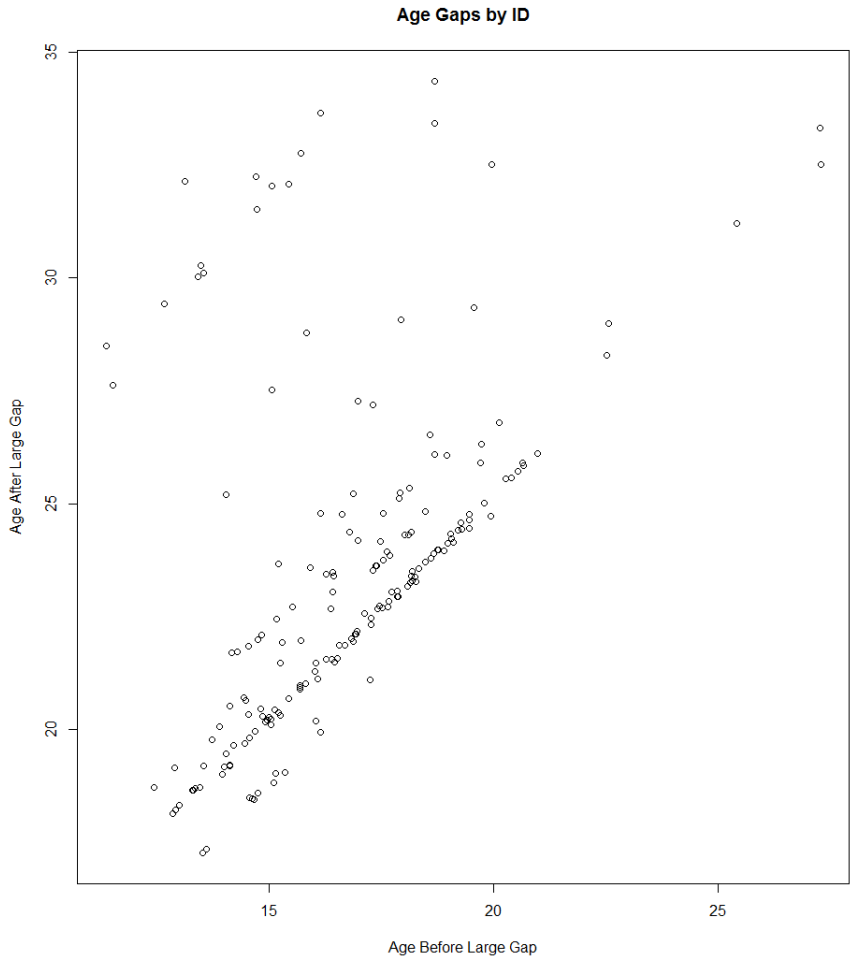
The ages and length of time encompassed in this second setting is highly variable. Some of these observations were removed from the data set. There were two main reasons for this. Most to all of bone growth is done by the time an individual reaches about 20 years of age. Almost all of the ages after the "break" were older than that. However, some of the ages were around 18-19. The idea was to not exclude all of these ages because it would give a better picture of adult TBBMC, which is what β is trying to measure. However, some of these observations were as high as 35-40. These are extreme outliers with respect to age. They were removed so that they would not influence the fit just based on how far away from the mean they were.

Next, was a decision on what would be an appropriate cut-off for age. The initial thought was about 20 years old, due to people's natural growth curves. Keeping this in mind, the data was used to find a natural break point around 20 years old. Figure E.7 best tells the story.

It is an attempt at age before and after the break for each individual. Granted, some of the 251 individuals do not have the second setting of observations. Looking at the graph, there is a clear gap around 20 to around 23 and then from around 23 to 25. Consequently, it was decided to eliminate all data with ages > 22. This eliminated most of the observations after the break, but still allowed for estimation of adult TBBMC since there were observations around 18-20 for most individuals. This eliminated about 28% of the data set (it went from 2,104 observations to 1,518). If the cut-off went up to 23, it would keep an additional 66 observations, or about 3% of the whole data set. If the cut-off went down to 20, it would lose an additional 108 observations, or about 5%. Another argument for cutting this at age 22, is that there is no discernible pattern for TBBMC vs. age after this point. This is illustrated in the 2-part graph labeled Figure E.8.

A few people's individual curves and data were examined. One person that had data of particular relevance was ID 1053. This person had quite a bit of data after

Figure E.7.: Age Before and After a Large Gap



22. In the Figure E.9, the black curve is their TBBMC values, while the blue and red lines represent ages of 20 and 22 respectively. This person exhibits only growth through age 22, but they have both growth and loss of TBBMC after this point. The goal is modeling adolescents, and their values should all be in the growth stage. Decline in TBBMC is another reason to remove high values of age.

Figure E.8.: Age vs. TBBMC before and after 22 years old

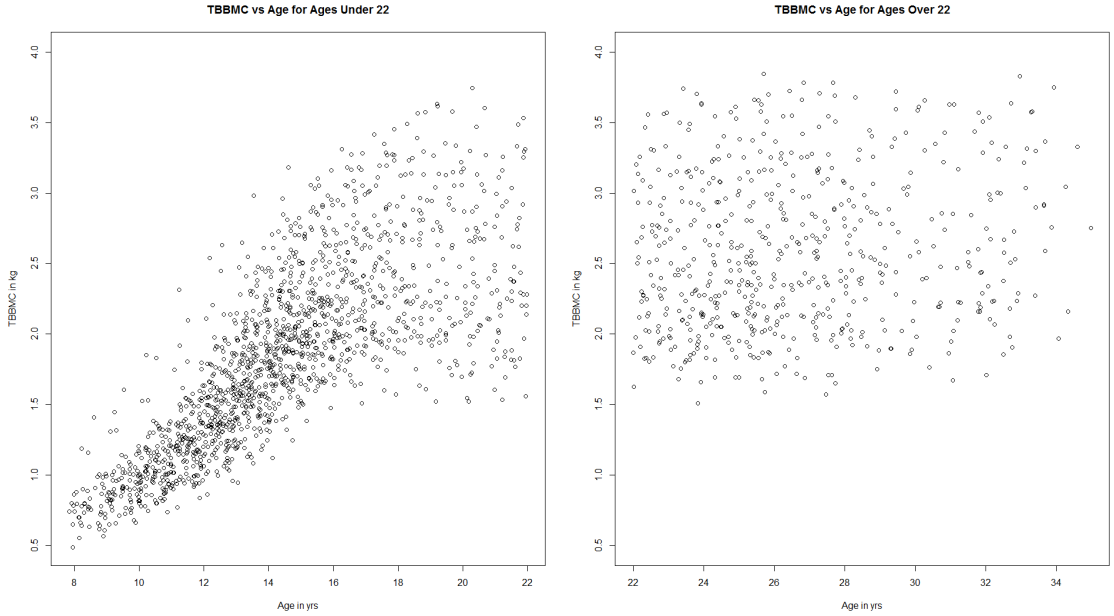
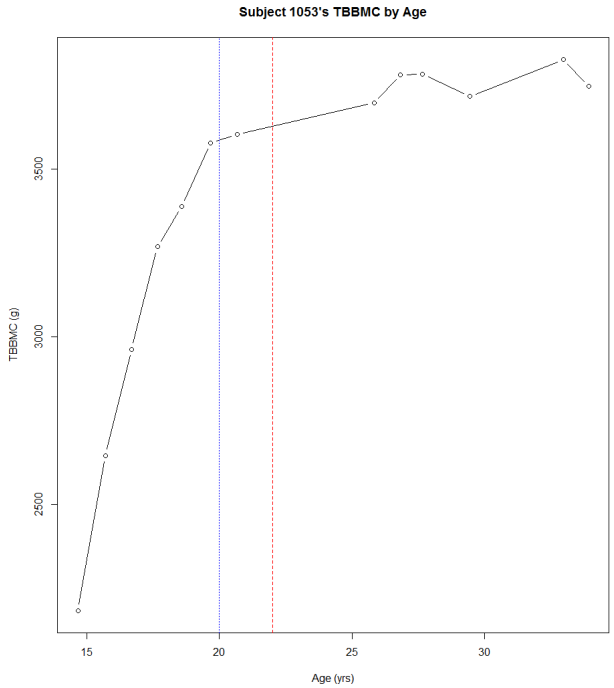


Figure E.9.: ID 1053's Age vs. TBBMC



Sensitivity Analysis

As stated previously, the data rules are for the individual parameterizations, and these are used to calculate starting values for the entire data set. The rationale for not eliminating the people who were not within the rules was that there was nothing wrong with their values, they just lacked clear peaks or data on both sides of the peak, so it was difficult to estimate their individual distributions. Here is a table of $\hat{\Theta}$ as well as the age and value of PBMCV for the overall model based on all of the data versus the data with the people outside the data rules ranges removed. This second data set is referred to as the Data Rules, to make the table easier to read.

Table E.2.: $\hat{\Theta}$ for all data versus within the data rules people

data set	All Data		Data Rules		Difference	
	Males	Females	Males	Females	Males	Females
n	692	826	614	759	82	67
μ	15.41	14.03	15.51	14.00	-0.09	0.03
β	2.78	2.14	2.81	2.14	-0.04	0.00
σ^{-1}	0.96	0.96	0.94	0.97	0.02	-0.01
Age of PBMCV (in y)	13.74	12.35	13.80	12.34	-0.06	0.02
PBMCV (in g/y)	310.34	238.98	308.80	241.62	1.54	-2.64

There is little difference in the estimates of the parameters or the important values of PBMCV and age of PBMCV whether the people for whom it was difficult to estimate a reasonable individual parameterization were included or excluded. This is one of the benefits of using this method. Before, Bailey et al. excluded people who did not have individual estimates for age and value of PHV and PBMCV. [4]

Appendix F: SAS

This is an example of code used to get individual parameters for each ID according to the generalized logistic regression model. In this particular example, since we doing it for each ID, there is no need to separate the parameters by gender. The model is $TBBMC = \beta * (1 + exp(-\sigma^{-1} * (Age - \mu)))^{-.2}$. Here TBBMC is measured in kg. In the code, BA is used for β, Mu, for μ, SI for σ^{-1}, F for TBBMC, and Age for Age.

One of the reasons to find the individual parameters is to test the joint normality. It should be noted that these output data sets were saved to be able to run any additional tests or plots on them. The data sets BParams and BParams2 have our individual parameters. The univariate procedure provides summary statistics, stem and leaf plots, boxplots, and a Normal QQ plot. These can be used to examine the individual distributions and normality of the parameters. Other code was used to assess joint normality, including tests such as Mardia, and Heinze-Zeinkler.

One of the last things in this code is a residual analysis. We let F = TBBMC. F is assumed normal with a predicted mean as the above equation, with some positive variance. This part is checking this assumption. The data set b2, looks at the residuals, $resid = F - \hat{F}$, as well as the squared residuals.

One drawback to this particular code, is that it does not have in a part about joint normality of the parameters in it. Since we are looking at individuals, there would be no way to assess it.

```
/* theta 5 individual parms */
proc nlmixed data=a1 maxiter=500; by ID;
   parms Mu=15 SI=.7 BA=2.6;
   F=BA*((1+exp(-SI*(Age-Mu)))**-.2);
   model TBBMCkg ~ Normal(F,s2);
   predict F out=b1;
   predict Mu out=bMu;
```

```
   predict SI out=bSI;
   predict BA out=bBA;
run;
data bMu; set bMu; Mu=pred;
data bSI; set bSI; SI=pred;
data bBA; set bBA; BA=pred;
data bParams; merge bBA bMu bSI;
proc sort data=BParams; by ID;
data BParams2; set BParams; by ID; if first.id;
run;

/* checking some assumptions with the parameters */
proc corr data=bParams2 cov; var Mu SI BA; run;
proc univariate data=bParams2 plot; var Mu SI BA; run;
proc gplot data=bParams2; plot (Mu Sigma SI)*BA/frame; run;
run;
/* calculating the residuals and checking the assumptions here */
data b2; set b1; resid=tbbmckg-pred; resid2=resid*resid;
proc univariate data=b2 plot; var resid resid2; run;
proc print data=b1; run;



proc export data=BParams2
 outfile='\\Client\E$\theta5parms.csv'
   dbms=csv
   replace;
run;
```

The point of this code is to highlight the previous drawback. This example uses $\Theta = (\mu, \sigma^{-1}, \beta, \gamma = .2)$. Since we are no longing examining them by ID, we can restrict the parameters to have joint normality directly in the code. Additionally, this will give us new parameters relating to the covariance matrices of the parameters. The initial guesses were found using the averages and covariances of the individual subjects' parameters.

```
/* theta 5 gender stratified model */
proc nlmixed data=a3 Method=FIRO qpoints=500 maxiter=500;
   parms MuMean=14.24 BMean=2.198 SIMean=1.207 MuMMean = 16.20
      BMMean = 2.953 SIMMean=1.256 s2=.00262
         s11=4.423966 s21=-.07516 s22=.11447 s31=-.15851 s32=-.06939
             s33=.822613
     s11m=2.900997 s21m=.070785 s22m=.180637 s31m=.376413 s32m
        =-.04326 s33m=.728731;
     if SexMale = 0 then
    F=B*((1+exp(-SI*(Age-Mu)))**-.2);
  if SexMale = 1 then
    F=BM*((1+exp(-SIM*(Age-MuM)))**-.2);
random Mu B SI MuM BM SIM ~
          Normal([MuMean,BMean,SIMean,MuMMean,BMMean,SIMMean],[s11,
             s21,s22,s31,s32,s33,0,0,0,s11m,0,0,0,s21m,s22m,0,0,0,
             s31m,s32m,s33m]) Subject=ID;
model TBBMCkg ~ Normal(F,s2);
estimate "Age of Peak Male" log(.2)/SIMMean+MuMMean;
estimate "Age of Peak Female" log(.2)/SIMean+MuMean;
estimate "Value of Peak Male" BMMean*SIMMean*((1+1/.2)**(-1.2));
estimate "Value of Peak Female" BMean*SIMean*((1+1/.2)**(-1.2));
run;
```

This particular code uses method=firo. There are 4 available methods within proc nlmixed, each with its own advantages and disadvantages. The other 3 methods are gauss, hardy, and isamp. Hardy is never used in this research because it is for one dimensional models, when we have either 3 or 4 dimensions. The other 2 methods were used at various points to compare results.

VITA

VITA

Michael Lawlor was born on October 14, 1983 in Oak Lawn, Illinois. He received his B.A. in Mathematics and Spanish from Wabash College in Crawfordsville, Indiana in 2006. Michael joined the Department of Statistics at Purdue University in 2007. He received his Master's degree in Statistics with an emphasis on Computational Finance in May 2009 and transitioned into the Ph.D. program at Purdue University.